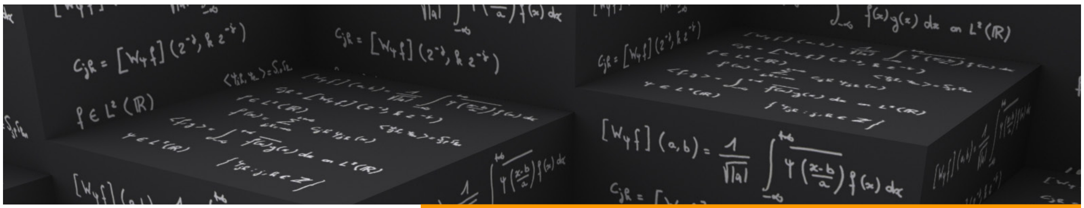


VIVIENNE
INVESTISSEMENT

Vivienne investissement

« Quantitative processes for active management »

Research report extracted from the article
« **Covariance versus Precision Matrix Estimation for Efficient Asset Allocation** »

published in the Special Issue of September 2016
« **Financial Signal Processing and Machine Learning for Electronic Trading** »
of the IEEE Journal on Selected Topics in Signal Processing

Laurent Jaffrès — CEO & Fund Manager
+33 4 78 60 50 58
ljaffres@vivienne-im.com

VIVIENNE INVESTISSEMENT
47 rue de la République - 69002 LYON
- FRANCE -
AMF : GP10000029 | 484 288 428 RCS Lyon

Covariance versus Precision Matrix Estimation for Efficient Asset Allocation

M. Senneret*, Y. Malevergne†, P. Abry‡, G. Perrin* and L. Jaffrès*
* Vivienne Investissement, Lyon, France
† Univ Lyon, UJM-Saint-Etienne, COACTIS, EA 4161, F-42023 Saint Etienne, France
‡ Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, F-69342 Lyon, France

***Abstract*—[1] Asset allocation constitutes one of the most crucial and most challenging tasks in financial engineering, which often requires the estimation of large covariance or precision matrices, from short time span multivariate observations, a mandatory yet difficult step. The present contribution reviews and compares a large selection of estimators for covariance and precision matrices, organized into classes of estimation principles (Direct, Factor, Shrinkage, Sparsity). This includes the theoretical derivation of several additional estimators not available in the literature. Rather than assessing estimation performance from synthetic data based on *a priori* selected models of questionable practical interest, it is chosen here to evaluate practically the *quality* of these estimators directly from portfolio selection performance, quantified by financial criteria. Portfolio selection is conducted over two datasets of different natures: a 15-year large subset (within Stoxx Europe 600) of 244 European stock returns, and a 50-year benchmark dataset of 90 US equity portfolios. This large scale comparative study addresses issues such as the relative benefits and difficulties of using robust versus direct estimates, of choosing precision or covariance estimates, of quantifying the impacts of constraints.**

***Index Terms*—Portfolio selection, asset allocation, covariance matrix, precision matrix, multivariate estimation, factor, shrinkage, sparsity.**

I. Introduction

Optimal asset allocation. Optimal asset allocation, i.e., the selection of a limited number of assets within a pre-selected (large-size) basket for optimal performance, constitutes a crucial stake in the finance industry [1], [21]. Amongst the numerous available strategies, the mean-variance framework of Markowitz [40] and in particular the global minimum variance portfolio (GMVP), is commonly used for actual asset allocation: Given a set of p risky securities, whose returns at time t are denoted $\{r_k(t), k = 1, \ldots, p, \, t = 1, \ldots, n\}$, the GMVP approach amounts to estimating the vector of weights $\{w_k, k = 1, \ldots, p\}$, that minimizes the overall standard deviation (referred to as the *volatility*) of the portfolio return:[2]

$$r(t) = \sum_{k=1}^{p} w_k \cdot r_k(t). \tag{1}$$

Without restriction on short-sales, i.e., weights w_k can be either positive or negative, the GMVP is obtained as the solution to the following optimization program:

$$w^* = \operatorname*{argmin}_w w'\Sigma w, \; s.t. \; 1_p' w = 1, \quad 1_p' := \underbrace{[1, \ldots, 1]}_{p-times}, \tag{2}$$

where $\cdot'$ denote the transpose operator and Σ the (full-rank) $p \times p$ covariance matrix of the asset returns. The well-known closed-form solution to problem (2) involves the $p \times p$ precision matrix, $\Pi := \Sigma^{-1}$, of the asset returns (e.g., [1]):

$$w^* = \left(1_p' \Pi \, 1_p\right)^{-1} \Pi \, 1_p. \tag{3}$$

This shows that the estimation of Σ, or more accurately of its inverse Π, constitutes a crucial step for a successful implementation of asset allocation, as well-known and abundantly documented [20], [15]. However the estimation of large covariance matrices is a notoriously difficult task: First, while pointwise convergence of usual covariance estimators is guaranteed under mild assumptions, often met in real-world data, estimation variance implies significant uncertainties and dispersion on eigenvalues and eigenvectors and thus the manipulation of ill-conditioned matrices; Second, large matrix inversion is a tedious, numerically unstable task, notably for ill-conditioned matrices (as often the case in practice, when the number of observations n is not large compared with the number of assets p). These issues are known to significantly impair the implementation of GMVP strategies [41], to the point that they may underperform equally weighted portfolio (EWP) strategies, that simply ignore asset dependence structure (i.e., $w_k = 1/p, \forall k$): For monthly return based US stock market, it was estimated that GMVP requires sample size n larger than $6,000$ months (500 years !) for a portfolio with $p = 50$ assets, in order to outperform EWP. This leads the authors of [15] to pessimistically conclude that *there are still many miles to go before the gains promised by optimal portfolio choice can actually be realized out-of-sample*[3].

Related work. A large amount of research efforts has been dedicated to overcoming the alternative brute force solution consisting of making use of the sole individual asset volatility, thus simply neglecting all dependencies amongst assets. It was notably reported that, compared with EWP or with GMVP based on the sample covariance matrix S_n, the *out-of-sample* volatility of GMVP can be significantly lowered by shrinking

[2]Given the prices $P_k(t)$ of asset k at time t, the corresponding return reads: $r_k(t) = (P_k(t+1) - P_k(t))/P_k(t)$.

[3]In the finance literature and econometrics, *out-of-sample* refers to performance computed on data that were not used for parameter estimation, tuning or learning.

S_n [36], [16], while the details of the shrinkage themselves have little impact [18]. However, despite reduced volatility, the benefits of shrinkage in terms of performance measured by financial criteria such as the Sharpe ratio (i.e., the ratio Gain/Volatility) are still questioned (cf. e.g., [32]).

Along another line, one can surprisingly note that most research efforts were dedicated to improving the estimation of Σ, while neglecting the obviously key matrix Π. It is however well-known that the inverse of S_n yields a poor estimate of the precision matrix, notably when $n \simeq p$ [42]. Amongst noticeable exceptions addressing directly the estimation of Π, one can refer to [33], [16] using shrinkage, and to [27] relying on sparsity and l_1-optimization. Further, restrictions on short-sales (i.e., additional constraints $w_k \geq 0, \forall k$) were reported to improve GMVP performance, essentially by acting as a shrinkage of the estimated Σ, and thus to a better conditioning of the *effective* covariance matrix [32]. Impact on direct estimates of Π remains yet undocumented.

Goals, contributions and outline. The overall goals of the present work are to contribute to a large size study of covariance/precision matrix estimation strategies in asset allocation and to compare their performances, measured with financial metrics and in realistic contexts. The following issues will notably be addressed: i) Should a direct estimate of the precision matrix be preferred to the inversion of an estimated covariance? ii) Which class of estimates of Σ, of Π should be preferred? in what context? iii) How do short sale restrictions or other additional financial constraints impact performance? To achieve such goals, the first originality of the present work is i) to review and organize existing estimators into classes of estimation principles (direct, factors, shrinkage and sparsity), and ii) to derive closed-form expressions for several reliable estimators for Σ and Π that had never been explicitly derived so far (cf. Section II). The second originality lies in performance assessment: Performance will not be quantified in terms of Mean-Squared-Error, or other classical signal processing estimation criteria, measured from average across Monte Carlo simulations conducted on synthetic and controlled data, as this would imply the recourse to a reliable multivariate statistical model for risky assets, for which a consensus is far from existing. Instead, performance will be evaluated using several financial performance assessment metrics (portfolio volatility, Sharpe ratio, asset turn over, Herfindal index...), computed by applying the full collection of devised estimators to two datasets of very different natures: i) a 15-year long large ($p = 244$) subset (within the Stoxx Europe 600 index) of actual risky assets, which constitute a relevant basket that can be used in practice by investment firms, hereafter denoted by Dataset I; ii) a 50-year long medium size ($p = 90$) dataset chosen because considered as a benchmark data set in the literature [16], [27], [23], hereafter denoted by Dataset II. Several realistic scenarios will be explored, permitting the analysis of the impacts of short sale restrictions on the estimates of both Σ and Π, as well as those of the ratio p/n quantifying estimation difficulty. Explorations are conducted up to $p/n = 3$, where estimation issues are significant, which has, to the best of our knowledge, not been addressed in the literature. Setting and performance are reported and commented in Section III. Future works and conclusions are envisaged in Section IV.

II. Covariance and precision estimation strategies

This section organizes the estimators for Σ and Π retained in this study, into four classes of statistical principles (direct, factor based, shrinkage based, sparsity based), and provides an overview of their main properties. It also proposes original derivations of several estimators for both the covariance and precision matrices.

A. Direct sample estimates

The sample covariance matrix S_n is the standard estimator,

$$\hat{\Sigma}^{(1)} := S_n \tag{4}$$

with $\mathrm{S_n(k,l)} =$

$$\frac{1}{n-1}\left[\sum_{t=1}^{n} r_k(t)\cdot r_l(t') - \frac{1}{n}\left(\sum_{t=1}^{n} r_k(t)\right)\cdot\left(\sum_{t=1}^{n} r_l(t')\right)\right], \tag{5}$$

However, it suffers from two main deficiencies. First, when $p/n > 1$, $\hat{\Sigma}^{(1)}$ is not full rank, and hence not invertible, thus requiring the use of the Moore-Penrose generalized inverse, denoted $\cdot^{+}$, to estimate the precision matrix:

$$\hat{\Pi}^{(1)} := \left[\hat{\Sigma}^{(1)}\right]^{+}. \tag{6}$$

Second, even when $\hat{\Sigma}^{(1)}$ is full rank, its inverse only provides a biased estimator of Π.

Efficient alternatives were considered, consisting of neglecting all dependencies amongst assets. The simplest replaces S_n with a scalar – actually the identity – matrix:

$$\hat{\Sigma}^{(2)} := \mathrm{Id}_p \quad \text{and} \quad \hat{\Pi}^{(2)} := \mathrm{Id}_p\,, \tag{7}$$

in which case GMVP simplifies to EWP. The diagonal matrix of the sample variances (i.e., the volatilities of each asset independently) is another straightforward possibility:

$$\hat{\Sigma}^{(3)} := \mathrm{diag}\left(\hat{\Sigma}^{(1)}\right) \quad \text{and} \quad \hat{\Pi}^{(3)} := \left[\hat{\Sigma}^{(3)}\right]^{-1}. \tag{8}$$

The use of $\hat{\Sigma}^{(2)}$ or $\hat{\Sigma}^{(3)}$ yields necessarily positive weights $w_k \geq 0, \forall k$, leading *de facto* to a so-called *long-only* strategy (no short sales).

The robust estimates detailed below constitute tentative trade-offs between estimates of the full dependence structure which show poor performance versus estimates that discard all dependencies. These direct estimates will thus provide benchmark strategies for the horse race in section III.

B. Factor

1) Factor principle: In order (i) to reduce the variance of estimator S_n, (ii) to ensure full rank estimates of Σ even when $p/n > 1$ and (iii) to get a reliable estimate of Π, the factor models provide a versatile approach. The simplest case is derived from the Sharpe market model [47] where the return of the market portfolio is assumed to be the single relevant factor. More general models based on three or four factors can provide better estimations for Σ specifically in the context of Dataset II [9], [22]. In this case the factors are taken as *exogenous*. Alternatively, when factors are unknown or unobservable, an approximate factor structure can be reconstructed from a principal component (or singular value) analysis [10]. Then, the factors are *endogenously* fixed.

2) Exogenous factors: For simplicity, we consider the single index market model as representative of the class endogenous factor models:

$$r_k(t) = \alpha_k + \beta_k \cdot r_m(t) + \varepsilon_k(t), \tag{9}$$

where $r_m(t)$ denotes the return on a market index representative of the class of assets under consideration, e.g. the Stoxx Europe 600 index for Dataset I and the S&P 500 for dataset II, on day t. Parameters $\alpha = (\alpha_k)_{k=1}^p$ and $\beta = (\beta_k)_{k=1}^p$ are the p-vectors of intercepts and factor loadings while $\varepsilon(t) = (\varepsilon_k(t))_{k=1}^p$ is the p-vector of idiosyncratic residuals with diagonal covariance matrix Δ. For simplicity, let us assume iid normal residuals. The covariance and precision matrices then read:

$$\Sigma = \sigma_m^2 \beta\beta' + \Delta \quad \text{and} \quad \Pi = \Delta^{-1} - \frac{\Delta^{-1}\beta \cdot \beta'\Delta^{-1}}{\frac{1}{\sigma_m^2} + \beta'\Delta^{-1}\beta}, \tag{10}$$

where $\sigma_m^2 = \mathrm{Var}\ r_m$.

Plug-in estimators can be obtained by naive replacement of β and Δ by their Ordinary Least Square (OLS) unbiased estimators. However, such plug-in estimators do not yield unbiased estimators for Σ and Π. Instead, taking into account the independence of the OLS estimators of β and Δ under normality and denoting by $\hat{\sigma}_m^2$ the unbiased estimate of σ_m^2, we derive:

$$\mathrm{E}\left[\hat{\sigma}_m^2 \hat{\beta}\hat{\beta}'\right] = \sigma_m^2 \beta\beta' + \frac{1}{n-1} \cdot \Delta\,, \tag{11}$$

which shows that the plug-in estimator of Σ exhibits a finite sample bias equal to $(n-1)^{-1}\Delta$. A bias-corrected estimator of the covariance matrix reads:

$$\hat{\Sigma}^{(4)} := \hat{\sigma}_m^2 \hat{\beta}\hat{\beta}' + \frac{n-2}{n-1} \cdot \hat{\Delta}\,, \tag{12}$$

which shows that $\hat{\Sigma}^{(4)}$ only differs from the plug-in estimator by the multiplicative factor $(n-2)/(n-1)$, stemming from the fact that the entries of $\hat{\Delta}$ are χ^2-distributed with $n-2$ degrees of freedom. As a consequence, this result generalizes to the case of a l-factor model by replacement of the multiplicative factor $(n-2)/(n-1)$ by $(n-l-1)/(n-1)$ in Eq.(12).

We can also propose an alternative estimator to the plug-in estimator for Π. To correctly compute (3), it is enough to obtain an estimator of Π up to a multiplicative constant, hence of the numerator of Π:

$$\left(1 + \sigma_m^2 \cdot \beta'\Delta^{-1}\beta\right) \cdot \Delta^{-1} - \sigma_m^2 \cdot \Delta^{-1}\beta \cdot \beta'\Delta^{-1}\,. \tag{13}$$

Taking into account the independence of the OLS estimators of β and Δ under normality, we can explicitly show that:

$$\mathrm{E}\left[\hat{\sigma}_m^2 \left(\hat{\beta}'\hat{\Delta}^{-1}\hat{\beta}\right)\hat{\Delta}^{-1} - \hat{\sigma}_m^2 \hat{\Delta}^{-1}\hat{\beta}\hat{\beta}'\hat{\Delta}^{-1}\right] = \left(\frac{n-2}{n-4}\right)^2 \cdot \begin{pmatrix} \sigma_m^2 \left(\beta'\Delta^{-1}\beta\right)\Delta^{-1} \\ \quad - \sigma_m^2 \Delta^{-1}\beta\beta'\Delta^{-1} + \frac{p-1}{n-1}\Delta^{-1} \end{pmatrix}, \tag{14}$$

and

$$\mathrm{E}\left[\hat{\sigma}_m^2 \cdot \hat{\beta}'\hat{\Delta}^{-1}\hat{\beta}\right] = \frac{n-2}{n-4}\left[\sigma_m^2 \cdot \beta'\Delta^{-1}\beta + \frac{p}{n-1}\right]. \tag{15}$$

Hence a suitable estimator of the precision matrix reads:

$$\hat{\Pi}^{(4)} := \frac{n-4}{n-2} \cdot \hat{\Delta}^{-1} - \frac{\left(\frac{n-4}{n-2}\right)^2 \hat{\sigma}_m^2 \hat{\Delta}^{-1}\hat{\beta}\hat{\beta}'\hat{\Delta}^{-1} - \frac{\hat{\Delta}^{-1}}{n-1}}{1 + \frac{n-4}{n-2} \cdot \hat{\sigma}_m^2 \hat{\beta}'\hat{\Delta}^{-1}\hat{\beta} - \frac{p}{n-1}}. \tag{16}$$

In the limit $n \to \infty$, with p fixed, $\hat{\Pi}^{(4)}$ converges towards the plug-in estimator. But, when both n and p grow so that $\lim_{n\to\infty} p(n)/n \to \gamma > 0$, in difference to the case of the covariance matrix, the correction in the denominator does not vanish and $\hat{\Pi}^{(4)}$ does not converge to the plug-in estimator.

3) Endogenous factors: When the factors are unobservable, one has to rely on Principal Component Analysis (PCA) or Singular Value Decomposition (SVD), from which many different estimators can be derived. We define two such pairs. For consistency with the previous approaches, we consider $\left(\hat{\Sigma}^{(5)}, \hat{\Pi}^{(5)}\right)$ whose definition follows that of $\left(\hat{\Sigma}^{(4)}, \hat{\Pi}^{(4)}\right)$ with the return on the market portfolio r_m replaced by the first SVD factor, extracted from the $p \times n$ matrix which stacks the time series of the returns r_k.

We further define $\hat{\Sigma}^{(6)}$, obtained from $\hat{\Sigma}^{(4)}$, using the two first factors of the SVD. However, we are so far not able to derive theoretically a two-factor estimator for Π, $\hat{\Pi}^{(6)}$ will thus not be considered.

C. The shrinkage approach

1) Shrinkage principle: Shrinkage was originally introduced in [49] to replace the full dependence estimator S_n by an optimal mix between S_n and a target, user supplied, matrix M (e.g., the choice $M = \hat{\Sigma}^{(3)} = \mathrm{Diag}\, S_n$ corresponds to mixing full dependence structure with the absence of any dependence):

$$\hat{\Sigma}^{shrink} = (1-\rho)\, S_n + \rho\, M. \tag{17}$$

Recently, [36] considered shrinkage toward a scalar matrix and toward the covariance matrix implied by Sharpe's market model (9) while [2] considered shrinkage toward a covariance matrix derived from a latent factor model estimated by principal component analysis. Further, [18] suggests that the simplest approach to shrinkage provides the best results. However, recent advances show that better approximations of the covariance matrix are achieved when the fat-tailed nature of data is accounted for [11], [12]. Non-linear shrinkages either based on an upper limit of the condition number of the estimated covariance matrix [51], [52] or on the Marcenko-Pastur equation [39] are shown to provide significant improvements [37]. The case of linear shrinkage is considered here only.

In principle, the optimal shrinkage parameter ρ is obtained from the minimization of the quadratic loss function, quantifying the bias/variance trade-off:

$$L(\rho, M) = \mathrm{E}\left[||\hat{\Sigma}^{shrink} - \Sigma||^2\right]. \tag{18}$$

The shrinkage parameter ρ for several classical models can be found in [46]. Various other merit functions were considered to derive optimal ρs, as alternatives to the criterion in Eq. 18[16].

2) Covariance matrix: Most approaches to shrinkage rely on criterion (18) to select the optimal ρ, and plug-in estimates replacing ensemble average with sample moments are commonly used (cf. Eq. 32 in Appendix A and [16]). Instead, we make use of the Oracle Approximating Shrinkage (OAS) estimator, introduced in [11] (for the case of the shrinkage toward the identity matrix), because its performance are actually close to those achieved with the oracle shrinkage estimator whose implementation would require the knowledge of the true Σ.

Lemma 1 (Chen *et al.* 2010, Theorem 3)**.** *Under the assumption of iid normally distributed asset returns, given the unbiasedness of S_n, the OAS estimator of S_n toward a scaled – by the average across assets volatility – identity matrix, reads:*[4]

$$\hat{\Sigma}^{(7)} = \hat{\rho} \cdot \frac{\operatorname{Tr} S_n}{p} \cdot \mathrm{Id}_p + (1-\hat{\rho}) \cdot S_n, \tag{19a}$$

$$\hat{\rho} = \min \left\{ \frac{\left(1-\frac{2}{p}\right) \operatorname{Tr}\left(S_n^2\right) + \left(\operatorname{Tr} S_n\right)^2}{\left(n-\frac{2}{p}\right) \cdot \left[\operatorname{Tr}\left(S_n^2\right) - \frac{\left(\operatorname{Tr} S_n\right)^2}{p}\right]}, 1 \right\}. \tag{19b}$$

When $n \to \infty$, $\hat{\rho} \to 0$ showing that S_n does not need improvement by shrinkage. Conversely, for small n, the leftmost term within the curl-brackets in (19b) can be larger than one, indicating that S_n cannot be reliably used.

We propose here to extend the dynamical system iterative procedure underlying the derivation of OAS to the shrinkage of S_n toward the diagonal Volatility matrix, which, to the best of our knowledge, has so far never been achieved (detailed derivation is given in Appendix A, Eqs. 33-40):

Lemma 2. *Under the assumption of lemma 1, the OAS estimator for Σ toward a diagonal matrix (consisting of the asset volatilities) reads:*

$$\hat{\Sigma}^{(8)} = \hat{\rho} \cdot \mathrm{Diag} S_n + (1-\hat{\rho}) \cdot S_n, \tag{20a}$$

$$\hat{\rho} = \min \left\{ \frac{\operatorname{Tr}\left(S_n^2\right) + \left(\operatorname{Tr} S_n\right)^2 - 2\operatorname{Tr}\left[\left(\mathrm{Diag} S_n\right)^2\right]}{n \cdot \left(\operatorname{Tr}\left(S_n^2\right) - \operatorname{Tr}\left[\left(\mathrm{Diag} S_n\right)^2\right]\right)}, 1 \right\}. \tag{20b}$$

3) Precision matrix: The shrinkage approach can also be successfully applied to the estimation of the precision matrix Π. When S_n is well-conditioned, i.e., when $p/n < 1$, [28] provides several random shrinkage estimators that outperform the naive estimator $\hat{\Pi} = (S_n)^{-1}$. The proposed strategy is, in essence, quite close to the strategy applied in [36] for the shrinkage of Σ. When S_n is singular, [34] recently provides a shrinkage method to improve on the classical Moore-Penrose generalized inverse.

In the context of portfolio optimization, the shrinkage of Π has only been recently considered in [33], [16], often with recourse to non-parametric cross-validation for the estimation of the shrinkage parameter. We depart from such approaches and elaborate on the dynamical system technology to explicitly derive two closed-form OAS estimator for Π when S_n is well-conditioned and its inverse admits a finite second order moment, i.e., when $n > p + 4$:

Lemma 3. *Under the assumption of iid normally distributed asset returns, given the unbiased sample precision matrix estimator $P_n = \frac{n-p-2}{n-1} \cdot S_n^{-1}$ with finite second order moment, the OAS estimator for shrinking Π toward an identity (inverse averaged volatility) matrix, when $n > p+4$, reads:*

$$\hat{\Pi}^{(7)} = \hat{\rho} \cdot \frac{\operatorname{Tr} P_n}{p} \cdot \mathrm{Id} + (1-\hat{\rho}) \cdot P_n, \tag{21a}$$

$$\hat{\rho} = \min \left\{ \frac{\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1} \cdot \operatorname{Tr}\left(P_n^2\right) + \frac{n-p-2-\frac{2}{p}}{n-p-1} \cdot \left(\operatorname{Tr} P_n\right)^2}{\left[\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1} + n - p - 4\right] \times \left[\operatorname{Tr}\left(P_n^2\right) - \frac{\left(\operatorname{Tr} P_n\right)^2}{p}\right]}, 1 \right\}. \tag{21b}$$

Lemma 4. *The OAS estimator for shrinking Π toward the diagonal matrix of inverse volatilities when $n > p+4$ reads:*

$$\hat{\Pi}^{(8)} = \hat{\rho} \cdot \mathrm{Diag}\left(P_n\right) + (1-\hat{\rho}) \cdot P_n, \tag{22a}$$

$$\hat{\rho} = \min \left\{ \frac{2 \cdot \operatorname{Tr}\left(\mathrm{Diag}\left(P_n\right)^2\right) + \frac{n-p}{n-p-1} \cdot \operatorname{Tr}\left(P_n^2\right) + \frac{n-p-2}{n-p-1} \cdot \left(\operatorname{Tr} P_n\right)^2}{\left(\frac{n-p}{n-p-1} + n - p - 4\right) \times \left[\operatorname{Tr}\left(P_n^2\right) - \operatorname{Tr}\left(\mathrm{Diag}\left(P_n\right)^2\right)\right]}, 1 \right\} \tag{22b}$$

The original and technical proofs for these two lemma are fully detailed in Appendix B.

The case where S_n does not admit a regular inverse is more difficult to handle, as the sample precision matrix P_n is usually estimated using the Moore-Penrose generalized inverse, whose statistic follows the generalized inverse Wishart distribution [5]. To the best of our knowledge, the moments of that distribution do not admit known closed-form expressions (except for cross-sectionally uncorrelated returns) [13]. The derivation of the shrinkage estimators $\hat{\Pi}^{(7)}$ and $\hat{\Pi}^{(8)}$ in the singular case, $(n < p+4)$, is hence left to future work.

D. The sparsity approach

1) Sparsity principle: The celebrated principle of parsimony (Occam's razor) has also long ago been summoned for large covariance or precision matrices estimation [17], [14], [25], [6], [8], [38]. Such a calling upon parsimony in that context may be motivated either from an *a priori* sparse modeling or from estimation issues.

Assuming *a priori* sparse dependence models, i.e., the fact that, beyond diagonal terms, only a small (compared to $p(p-1)$) number of entries of Σ or Π theoretically differ from zero may first stem from some *theoretical* or *background* knowledge on the system governing the data at hand: Assets belonging to a given class shall be related

[4] This reproduces a result in [11] but for two corrections: i) $n \to n-1$ accounts for the expectation being actually unknown, ii) a recurrent typo in [11](III.C) requires to replace $\frac{n+1-2}{p}$ with $n+1-\frac{2}{p}$ and $\frac{1-2}{p}$ with $1-\frac{2}{p}$.

together while assets pertaining to different classes are more likely to be independent. It then remains an open and difficult question to decide whether such a relative *independence* of classes of assets is better modeled with non diagonal zeroed entries in the covariance or in the precision matrix. When the covariance matrix is chosen sparse, its corresponding inverse, the precision matrix, is usually not sparse (and vice-versa). As a consequence, assuming that either the covariance or the precision matrix is sparse amounts to choosing from the very beginning between two different structural models. Sparse covariance is equivalent, in a Gaussian framework, to consider that the corresponding covariates are independent. It is likely more relevant when one considers assets traded on different markets with weak cross-market correlations, thus yielding block-sparse covariance matrices. Conversely, sparse precision corresponds, within that same framework, to covariates that are *conditionally* independent. It thus appears more naturally when assets returns can be assumed to be linearly related, so that given the knowledge of a subset, the remainders are uncorrelated. Beyond these theoretical considerations, the numerical experimentations and analyses reported in Section III below can be read as elements of answers, in the context of practical portfolio allocation performance, to the challenging issue of deciding between a sparse structure *a priori* imposed to Σ or Π.

The second category of reasons motivating sparsity in dependence matrices stems from the well-known *screening* effect that accompanies large covariance or precision matrix estimation [30]: For large matrices, estimated from short sample size, i.e., when $n > p$ or even when $n \lesssim p$, estimation performance for the non diagonal entries are such that it cannot be decided whether small values correspond to actual non zero correlations or to estimation fluctuations, and thus noise. Therefore, small values should be discarded and large values only are significantly estimated and should be further used.

For both issues – sparse modeling or estimation issues – the practical challenge is to decide how many and which non diagonal entries should be set to zero. A natural way to enforce sparsity is to resort to direct thresholding of the non diagonal entries of the estimated matrices (cf. e.g., [4], [44], [7], [53]). While this received significant theoretical developments and laid to success in numerous applications, this yields covariance matrices that are not *a priori* positive definite for finite sample sizes [26], whose inverse thus does not correspond to a Precision matrix.[5] Experiments not reported here lead us to observe that such estimates yield poor asset allocation performance and thus to conclude that they are not suited for that particular purpose. Instead, sparsity is enforced by minimizing a cost function, balancing a data fidelity term and a penalty term based on the l_1-norm of the non diagonal entries of the estimated matrix, thus following the large bulk of ongoing efforts aiming to address sparse matrix estimation issues, cf. [17], [14], [25], [6], [8], [38] for Σ and more recently [3], [45] for Π.

[5]Consistency of threshold estimates have been studied in [4], which proves that consistency can be achieved when the true covariance matrix is sparse, the variables are Gaussian or sub-Gaussian, in the limit $(\log p)/n \to 0$.

2) Precision matrix: For Π, the state-of-the-art formulation of l_1-norm sparsity induced estimator is now referred to as the *Graphical Lasso* [25]: It balances the negative log-likelihood function, thus relying on the Graphical Gaussian Model framework, and hence following the original formulation due to [17], with an l_1 penalization of $\hat{\Pi}$:

$$\hat{\Pi}^{(9)} := \underset{\Pi}{\operatorname{argmin}} \operatorname{Tr}(S_n \Pi) - \log\det\Pi + \lambda \cdot ||\Pi||_1, \qquad (23)$$

where λ denotes a penalization parameter to be selected. Indeed, l_1 penalization has been observed to act as an efficient surrogate of l_0 penalization, that explicitly counts non zero entries, yet results in a non convex optimization problem. Instead, estimating Π from Eq. (23) thus amounts to solving a convex optimization problem, and practical solutions were described in the literature, the two most popular relying on the so-called *path-wise coordinate descent* [3] or *Alternating Direction Method of Multipliers* algorithms [6]. In the present contribution, use is made of this latter algorithm. Sparse estimators for Π were recently used in the context of portfolio selection in [27].

3) Covariance matrix: Sparsity can be imposed onto the covariance matrix through an equivalent formulation:

$$\hat{\Sigma}^{(9)} := \underset{\Sigma}{\operatorname{argmin}} \operatorname{Tr}(S_n \Sigma^{-1}) + \log\det\Sigma + \lambda \cdot ||\Sigma||_1, \qquad (24)$$

that however results in a non-convex problem and is hence far more difficult to solve. It has however been observed that the argument in Eq. (24) can actually be split into a concave and a convex function, and that minimization can thus be performed by a majorization-minimization algorithm [3]. In conjunction with a factor-based approach, sparse estimation of the error covariance matrix was recently applied in the context of portfolio selection in [23], [24].

III. COMPARED PORTFOLIO SELECTION PERFORMANCE

A. Datasets and performance criteria

1) Datasets: To review and compare performance in GMVP strategies, the 9 different estimators for Σ and Π, studied in Section II above, have been applied to two datasets of very different natures and sizes.

Dataset I is chosen as a realistic and typical real-world basket of assets portfolio managers could actually use. It consists of the daily returns of the $p = 244$ largest capitalizations of the Stoxx Europe 600 index, for a period of 15 years, from 2000, May 1st to 2015, August 31st, i.e., for $4'000$ trading days, thus constituting a realistic set of $244 \times 4'000 = 976'000$ observations.

Dataset II consists of the daily returns of the 100 value-weighted Fama and French US equity portfolios sorted by size and book-to-market value starting 1926, July 1st.[6,7] This dataset is chosen here because it is considered a benchmark in previous academic articles dedicated to covariance matrix estimation in a financial context (e.g., [16], [27]). To minimize

[6]Data available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/100_Portfolios_10x10_Daily_TXT.zip

[7]See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_100_port_sz.html for details on the portfolios construction

the impact of the significant amount of missing data, analysis is focused on $p = 90$ portfolios and on the time period from 29/04/1955 to 31/08/2015, i.e. $15'188$ trading days, which delivers a balanced panel of $90 \times 15'188 = 1'366'920$ observations.

2) Strategy set up: A crucial issue in portfolio allocation lies in stock market time series being highly non-stationary, thus naturally raising the question of assessing the typical *stationarity time scale*. This question is however ill-posed and its correct formulation requires the explicit and detailed formulation of the problem at stakes. In GMVP strategies, for the allocation a time t, the estimation of Σ and Π is conducted using a sliding winding along the past n days. The non-stationarity issue thus translates into selecting the size n that achieves optimal performance. The question needs to be further specified with respects to the size of the covariance matrix $p \times p$ and the chosen performance metrics. There has been several interesting attempts to tune automatically and adaptively the optimal sliding window size to the data, whose results yet remain difficult to interpret. The focus is thus here on the comparisons of the performance of the different estimators themselves, and not on the evaluation of the adaptive tuning strategies. Therefore, to avoid ambiguities in comparisons, performance are hence measured for different but pre-selected window sizes. Performance are evaluated for several ratios p/n, varied from $1/2 < 1$ (S_n is a full rank matrix) up to $3 > 1$ (S_n is highly singular but this corresponds to realistic portfolio selection operational frameworks), which goes far beyond previous studies. For Dataset I: $p/n = 1/2, 3/4, 1.1, 2$, corresponding to 24, 18, 12 and 6 months of trading days respectively. For Dataset II, $p/n = 3/4, 1, 2, 3$, corresponding to 6, 4.5, 3 and 1.5 months of trading days respectively. For both datasets, portfolios are rebalanced once a week (i.e, one day out of five). For ease of comparisons, transaction costs are not accounted for.

3) Parameter tuning: The three first classes (direct, factor, shrinkage) of proposed estimators do not imply any parameter tuning. The Sparsity estimator class does require the tuning of parameter λ. There has been several interesting work investigating automated and adaptive tunings of λ, e.g., [43]. Again to avoid blurring in performance comparisons stemming from the behavior of the tuning procedure itself, GMVP has been performed systematically for a large collection of *a priori* fixed λ, over the first half of Dataset I only. The same λ, without further tuning is used for the second half of Dataset I and for the full duration of Dataset II. Results are reported for λ yielding the best GMVP performance only, thus slightly favoring that class. It has been checked that this induces a mild bias only as the it was observed that performance vary only weakly and smoothly on a large range of values of λ.

4) Constraints from short sale restrictions: For realistic performance assessment in a real-world financial set up, we run simulation both with and without short sale restrictions. Forbidding short sells (also referred to as *long only* portfolios) means that the weights w_k are subjected to constraints of the form $w_k \geq 0$, that can be read as a generalized inequality over the nonnegative orthant:

$$w^* = \underset{w}{\operatorname{argmin}}\, w'\Sigma w, \ s.t. \ 1_p' w = 1, \text{ and } \quad w \geq 0, \tag{25}$$

Solving Eq. 25 requires the inversion of very large matrices consisting of the (large) covariance matrix further augmented with blocks accounting for the constraints (see Supplementary Material for further details). To address both issues of inverting such a large matrix and of accommodating estimates of either Σ or of Π, we have developed an original block inversion procedure. When estimators for Σ are considered, the (pseudo-) inverse of $\hat{\Sigma}$ is used. When estimators for Π are considered, estimates $\hat{\Pi}$ are directly used to replace the corresponding block inverse. Interested readers are further referred to Supplementary Material.

5) Minimization procedures: The minimization procedures have been designed by ourselves, both for the constrained and unconstrained cases. The sparse precision estimator $\hat{\Pi}^{(9)}$ in Eq. 23 has been implemented after the procedure described in [31]. We obtained the sparse covariance estimator $\hat{\Sigma}^{(9)}$ in Eq. 24 from [3], and the corresponding procedure has kindly been made available to us by the authors.

6) Performance assessment: Estimation performance are classically estimated in terms of Mean-Squared-Error (MSE), which however requires that ground truth and thus that estimation is applied to synthetic data, using *a priori* prescribed models. In stock market modeling, there is no general consensus on the validity of specific models, notably because of the stationarity issue discussed above. Therefore, the relevance of performance assessed on synthetic data remains always controversial with respect to their validity when applied to real financial data. To overcome this difficulty, it is chosen here to assess performance directly on financial data and to compare efficiency of the proposed estimators making use of well-accepted and practically meaningful financial performance assessment metrics: GMVP aims to minimize the standard deviation (volatility, V) of the achieved portfolio, this is hence a natural criterion for performance assessment ; Sharpe ratio (S), that balances the gain (average return) and the risk (average volatility), is also considered a crucial index in portfolio management; Asset turn over (TO),

$$\mathrm{TO} = \frac{1}{n-1} \sum_{t=1}^{n-1} \sum_{k=1}^{p} |w_k(t+1) - w_k(t)|, \tag{26}$$

is also critical as changes in allocation imply transaction costs and further issues in actually achieving the allocation (liquidity, order operations,...), a low turnover is hence preferred by practitioners ; The inverse of the Herfindal index (H^{-1}),

$$\mathrm{H}^{-1} = \left(\frac{1}{n} \sum_{t=1}^{n} \sum_{k=1}^{p} w_k^2(t) \right)^{-1}, \tag{27}$$

is used to measure the portfolio diversification (that is the larger H^{-1} the larger the number of assets the portfolio is actually invested on). This is another important index for risk assessment, because too concentrated portfolios are usually regarded as lacking robustness and often result in very unstable allocations which yield high (and, in practice, expensive) turn-over rates. Out-of-sample (also referred to as *ex-post*) performance only are reported.

B. Performance comparisons

For space reasons, portfolio selection performance restricted to two ratios n/p only per dataset are presented in the main body of the text: For Dataset I, $p/n = 2$ and 1, for Σ and Π in Tables I and II respectively ; For Dataset II, $p/n = 3$ and 1, for Σ and Π in Tables III and IV respectively. Empirical standard errors (that could serve for statistical pairwise comparisons) reported every second line complement empirical means, both obtained as average over the entire observation period. Portfolio selection performance are further reported exhaustively, and following a different organization, in Supplementary Material. Interested readers are referred to Supplementary Material for compared illustrations of estimated dependence matrices and portfolio Net Asset Values. In Tables II and IV, *N/A* is used to indicate when Precision estimators are not available.

1) Direct versus robust estimates: Focusing on the Sharpe ratio S as the major practical performance index, let us first consider the direct estimates. Estimate $\hat{\Sigma}^{(1)}$ aiming at the full dependence differs in behavior from $\hat{\Sigma}^{(2)}$ or $\hat{\Sigma}^{(3)}$ neglecting all dependencies: Under $\hat{\Sigma}^{(1)}$, GMVP allocation is achieved by a small number of long-short positions on *highly* correlated pairs of assets (low H^{-1}), pairs that however keep changing at each allocation (high TO), a hence very unstable and undesired behavior, while $\hat{\Sigma}^{(2)}$ or $\hat{\Sigma}^{(3)}$ yield constant investment on a large number of assets (high H^{-1}). Under short-sale restrictions, long-short positions on highly correlated pairs are no longer possible and $\hat{\Sigma}^{(1)}$ then follows the same behavior as $\hat{\Sigma}^{(2)}$ or $\hat{\Sigma}^{(3)}$. All together, performance yield by $\hat{\Sigma}^{(1)}$ are poorer than those achieved with $\hat{\Sigma}^{(3)}$, thus showing that a crude attempt to estimate the full dependence structure might be worse than neglecting all dependencies. Further, Tables I to IV, together with those reported in Supplementary Material, unambiguously show that direct estimates are significantly outperformed by robust estimates, irrespective of the robustness principle (factor, shrinkage, sparsity) and of the ratio p/n, thus showing the clear and significant need and benefit of robust estimates, as an optimal trade-off between a non robust estimation of the full dependence structure and neglecting all dependencies.

2) Within $\hat{\Sigma}$: For the estimators of Σ, Tables I and III show, consistently for both datasets, that Factor and Shrinkage both yield much improved performance, compared to direct estimates, while Sparsity is not operative in delivering improved performance, and this irrespective of the ratio p/n. Shrinkage-based and factor-based estimators are however not equivalent as, when n becomes large, shrinkage reduces volatility while factor estimates favor the Sharpe ratios. Furthermore, shrinkage, compared to factor, yields concentrated portfolios with high turn-overs. For both classes of estimators, diversification (H^{-1}) does not depend on p/n, while the turn-over decreases notably when p/n decreases. While the performance of Factor and shrinkage show overall comparable performance, Factor does slightly better for Dataset I while Shrinkage performs best for Dataset II. This difference can be ascribed to the stronger departure from the normality assumption for Dataset I relative to Dataset II. Indeed, Dataset I is made of individual stocks with fat-tail return distributions while Dataset II is comprised of portfolios, hence with return distributions closer to normality by virtue of the Central Limit Theorem.

3) Within $\hat{\Pi}$: For the estimators of Π, the first striking observation from both datasets lies on the impressive performance of sparsity: Volatility is remarkably low with a quite large Sharpe ratio, irrespective of p/n. This is in agreement with results reported in [27] on different datasets and settings. Second, Factor yields decent performance (yet lower than those achieved with Sparsity). Third, consistently for both datasets, shrinkage performs poorly. This likely stems from the choice made for the target matrix, as $\mathrm{Diag}\, P_n$. Indeed, the inversion of S_n to obtain P_n already yields significant estimation noise, that may have been avoided by considering $(\mathrm{Diag}\, S_n)^{-1}$ as a target matrix. Also, the departure from the

TABLE I
GMVP Performance – Dataset I: Stoxx Europe 600 – Σ.
V: volatility, S: Sharpe ratio, TO: Turn Over, H^{-1}: inverse Herfindal index (diversification). Top: Long/Short allocation, Bottom: Long Only. Best performers are highlighted in bold.

	$n = 125,\ p/n \simeq 2$				$n = 250,\ p/n \simeq 1$			
	V	S	TO	H^{-1}	V	S	TO	H^{-1}
	Σ - Long/Short							
	Direct							
$\hat{\Sigma}^{(1)}$	296.3	0.35	363.4	0.11	82.35	-0.02	48.42	0.15
	(39.74)	(0.74)	(1508.56)	(0.17)	(8.51)	(0.36)	(33.90)	(0.13)
$\hat{\Sigma}^{(2)}$	19.65	0.67	0.00	244.0	19.65	0.67	0.00	244.0
	(0.45)	(0.28)	(0.00)	(0.00)	(0.44)	(0.27)	(0.00)	(0.00)
$\hat{\Sigma}^{(3)}$	15.71	0.67	0.04	164.6	16.15	0.67	0.02	172.8
	(0.37)	(0.28)	(0.02)	(21.59)	(0.41)	(0.25)	(0.01)	(17.92)
	Factor							
$\hat{\Sigma}^{(4)}$	**9.31**	**1.19**	**0.19**	**29.78**	**10.02**	**1.18**	**0.11**	**28.24**
	(0.23)	(0.28)	(0.12)	(15.71)	(0.30)	(0.31)	(0.07)	(12.70)
$\hat{\Sigma}^{(5)}$	9.42	1.09	0.18	34.28	10.00	1.11	0.10	32.59
	(0.24)	(0.27)	(0.11)	(18.57)	(0.27)	(0.29)	(0.07)	(15.06)
$\hat{\Sigma}^{(6)}$	8.56	1.11	0.25	29.09	**9.05**	**1.22**	**0.15**	**26.20**
	(0.21)	(0.24)	(0.16)	(16.59)	(0.24)	(0.25)	(0.12)	(10.77)
	Shrinkage							
$\hat{\Sigma}^{(7)}$	**8.07**	**1.18**	**0.98**	**15.26**	8.63	0.91	1.12	8.12
	(0.19)	(0.24)	(0.38)	(8.93)	(0.18)	(0.25)	(0.47)	(4.32)
$\hat{\Sigma}^{(8)}$	7.43	1.14	0.97	10.23	8.15	0.85	1.06	6.64
	(0.20)	(0.29)	(0.38)	(5.72)	(0.19)	(0.27)	(0.43)	(3.28)
	Sparsity							
$\hat{\Sigma}^{(9)}$	249.7	0.20	134.5	0.05	16.15	0.67	0.02	172.8
	(19.26)	(0.62)	(98.48)	(0.08)	(0.38)	(0.23)	(0.01)	(17.92)
	Σ - Long Only							
	Direct							
$\hat{\Sigma}^{(1)}$	9.14	1.06	0.24	10.27	9.19	1.14	0.13	10.51
	(0.23)	(0.27)	(0.14)	(5.83)	(0.23)	(0.29)	(0.09)	(5.02)
$\hat{\Sigma}^{(2)}$	19.65	0.67	0.00	244.0	19.65	0.67	0.00	244.0
	(0.49)	(0.26)	(0.00)	(0.00)	(0.44)	(0.25)	(0.00)	(0.00)
$\hat{\Sigma}^{(3)}$	15.74	0.67	0.04	165.6	16.18	0.67	0.02	173.7
	(0.40)	(0.25)	(0.02)	(21.48)	(0.37)	(0.26)	(0.01)	(17.87)
	Factor							
$\hat{\Sigma}^{(4)}$	9.18	1.10	0.17	15.17	**9.25**	**1.21**	**0.10**	**14.09**
	(0.26)	(0.27)	(0.11)	(11.48)	(0.23)	(0.26)	(0.07)	(8.09)
$\hat{\Sigma}^{(5)}$	9.33	1.02	0.17	17.16	9.46	1.14	0.10	15.74
	(0.24)	(0.25)	(0.10)	(14.59)	(0.24)	(0.26)	(0.06)	(10.07)
$\hat{\Sigma}^{(6)}$	9.18	1.04	0.18	15.53	9.23	1.18	0.10	13.87
	(0.23)	(0.28)	(0.12)	(13.25)	(0.25)	(0.27)	(0.08)	(7.51)
	Shrinkage							
$\hat{\Sigma}^{(7)}$	**9.21**	**1.18**	**0.22**	**16.54**	**9.21**	**1.20**	**0.13**	**13.08**
	(0.22)	(0.24)	(0.12)	(10.62)	(0.20)	(0.27)	(0.08)	(6.32)
$\hat{\Sigma}^{(8)}$	9.10	1.08	0.23	11.50	9.18	1.15	0.13	11.08
	(0.24)	(0.23)	(0.13)	(7.13)	(0.22)	(0.29)	(0.08)	(5.42)
	Sparsity							
$\hat{\Sigma}^{(9)}$	12.41	0.85	0.30	68.88	16.18	0.67	0.02	173.7
	(0.33)	(0.26)	(0.06)	(20.15)	(0.43)	(0.28)	(0.01)	(17.87)

TABLE II
GMVP Performance – Dataset I: Stoxx Europe 600 – Π. Same legend as table I.

	$n=125$, $p/n \simeq 2$				$n=250$, $p/n \simeq 1$			
	V	S	TO	H^{-1}	V	S	TO	H^{-1}
Π - Long/Short								
Factor								
$\hat{\Pi}^{(4)}$	9.71	1.15	0.22	26.44	10.30	1.16	0.11	26.47
	(0.21)	(0.26)	(0.14)	(15.22)	(0.27)	(0.28)	(0.08)	(12.52)
$\hat{\Pi}^{(5)}$	85.77	0.44	3.07	64.57	10.09	1.10	0.10	33.00
	(17.75)	(0.31)	(16.80)	(66.63)	(0.29)	(0.24)	(0.07)	(14.59)
Shrinkage								
$\hat{\Pi}^{(7)}$	N/A	N/A	N/A	N/A	19.62	0.62	0.26	223.7
					(0.48)	(0.28)	(0.40)	(45.78)
$\hat{\Pi}^{(8)}$	N/A	N/A	N/A	N/A	18.39	0.56	0.74	87.94
					(0.50)	(0.28)	(0.30)	(25.69)
Sparsity								
$\hat{\Pi}^{(9)}$	**7.05**	**1.45**	**0.41**	**18.65**	**7.16**	**1.57**	**0.22**	**19.36**
	(0.22)	(0.27)	(0.16)	(8.75)	(0.18)	(0.28)	(0.11)	(7.96)
Π - Long Only								
Factor								
$\hat{\Pi}^{(4)}$	45.01	0.33	0.22	12.96	15.14	0.86	0.14	2.96
	(1.12)	(0.27)	(0.16)	(37.55)	(0.67)	(0.24)	(0.17)	(3.18)
$\hat{\Pi}^{(5)}$	35.30	0.29	0.12	80.67	9.94	1.07	0.10	14.85
	(0.96)	(0.24)	(0.10)	(53.95)	(0.28)	(0.26)	(0.07)	(11.78)
Shrinkage								
$\hat{\Pi}^{(7)}$	N/A	N/A	N/A	N/A	19.66	0.61	0.09	239.3
					(0.48)	(0.28)	(0.11)	(13.10)
$\hat{\Pi}^{(8)}$	N/A	N/A	N/A	N/A	18.51	0.55	0.60	104.5
					(0.48)	(0.26)	(0.14)	(24.00)
Sparsity								
$\hat{\Pi}^{(9)}$	**8.96**	**1.16**	**0.19**	**13.88**	**9.13**	**1.23**	**0.10**	**14.40**
	(0.22)	(0.25)	(0.11)	(8.67)	(0.26)	(0.31)	(0.07)	(7.27)

TABLE III
GMVP Performance – Dataset II: Fama-French Portfolios – Σ. Same legend as table I.

	$n=30$, $p/n \simeq 3$				$n=90$, $p/n \simeq 1$			
	V	S	TO	H^{-1}	V	S	TO	H^{-1}
Σ - Long/Short								
Direct								
$\hat{\Sigma}^{(1)}$	211.3	0.66	561.6	0.12	207.6	-1.22	150.9	0.03
	(47.20)	(1.04)	(2081.58)	(0.21)	(87.28)	(2.77)	(108.74)	(0.03)
$\hat{\Sigma}^{(2)}$	15.26	0.99	0.00	90.00	15.26	0.99	0.00	90.00
	(0.27)	(0.13)	(0.00)	(0.00)	(0.26)	(0.14)	(0.00)	(0.00)
$\hat{\Sigma}^{(3)}$	13.61	1.14	0.11	67.45	13.79	1.13	0.04	71.67
	(0.20)	(0.13)	(0.04)	(8.33)	(0.22)	(0.12)	(0.02)	(7.21)
Factor								
$\hat{\Sigma}^{(4)}$	8.45	2.20	0.84	8.06	8.79	2.20	0.35	6.63
	(0.25)	(0.14)	(0.45)	(4.62)	(0.30)	(0.16)	(0.24)	(3.47)
$\hat{\Sigma}^{(5)}$	7.95	2.30	0.87	7.90	8.30	2.35	0.37	6.20
	(0.21)	(0.17)	(0.44)	(5.22)	(0.25)	(0.17)	(0.26)	(2.83)
$\hat{\Sigma}^{(6)}$	7.23	2.59	1.08	6.79	7.39	2.58	0.50	5.25
	(0.11)	(0.14)	(0.51)	(3.78)	(0.13)	(0.14)	(0.34)	(2.32)
Shrinkage								
$\hat{\Sigma}^{(7)}$	7.03	2.64	1.70	6.36	7.13	2.60	1.82	2.40
	(0.14)	(0.15)	(0.64)	(4.91)	(0.12)	(0.15)	(0.68)	(1.41)
$\hat{\Sigma}^{(8)}$	**6.90**	**2.77**	**1.79**	**4.64**	**7.05**	**2.64**	**1.72**	**2.24**
	(0.15)	(0.15)	(0.64)	(3.19)	(0.17)	(0.18)	(0.68)	(1.13)
Sparsity								
$\hat{\Sigma}^{(9)}$	17.32	0.88	1.45	66.06	13.70	1.14	0.09	71.08
	(1.69)	(0.18)	(28.71)	(9.76)	(0.22)	(0.14)	(0.85)	(8.47)
Σ - Long Only								
Direct								
$\hat{\Sigma}^{(1)}$	9.45	2.09	0.69	4.58	9.34	2.16	0.26	4.88
	(0.14)	(0.15)	(0.36)	(2.39)	(0.16)	(0.14)	(0.18)	(2.64)
$\hat{\Sigma}^{(2)}$	15.26	0.99	0.00	90.00	15.26	0.99	0.00	90.00
	(0.22)	(0.14)	(0.00)	(0.00)	(0.25)	(0.14)	(0.00)	(0.00)
$\hat{\Sigma}^{(3)}$	13.62	1.14	0.11	67.79	13.80	1.13	0.04	71.89
	(0.25)	(0.13)	(0.04)	(8.21)	(0.19)	(0.12)	(0.02)	(7.12)
Factor								
$\hat{\Sigma}^{(4)}$	**9.46**	**2.23**	**0.49**	**7.17**	**9.52**	**2.31**	**0.18**	**6.74**
	(0.12)	(0.14)	(0.26)	(4.71)	(0.15)	(0.15)	(0.13)	(3.36)
$\hat{\Sigma}^{(5)}$	9.45	2.19	0.53	6.27	**9.47**	**2.29**	**0.20**	**5.67**
	(0.16)	(0.15)	(0.29)	(5.16)	(0.15)	(0.14)	(0.15)	(2.81)
$\hat{\Sigma}^{(6)}$	9.37	2.17	0.56	5.66	9.37	2.25	0.21	5.39
	(0.15)	(0.13)	(0.30)	(3.80)	(0.15)	(0.14)	(0.15)	(2.76)
Shrinkage								
$\hat{\Sigma}^{(7)}$	9.48	2.05	0.55	8.55	9.35	2.13	0.24	6.08
	(0.15)	(0.12)	(0.26)	(6.25)	(0.15)	(0.15)	(0.16)	(3.41)
$\hat{\Sigma}^{(8)}$	9.40	2.08	0.62	6.07	9.36	2.14	0.25	5.41
	(0.14)	(0.15)	(0.29)	(4.27)	(0.17)	(0.16)	(0.16)	(3.00)
Sparsity								
$\hat{\Sigma}^{(9)}$	13.47	1.16	0.12	66.86	13.73	1.14	0.04	71.36
	(0.22)	(0.13)	(0.06)	(8.88)	(0.25)	(0.14)	(0.04)	(7.89)

normality assumption may lower the efficiency of the AOS estimators.

4) Impact of short sale restrictions: Short sale restrictions impact portfolio selection in several ways, consistent for both datasets and irrespective of p/n. i) The overall performance obtained from all estimators are consistently degraded in terms of Sharpe ratio and show lower turn over and larger diversification (lower H^{-1}). ii) The differences between the performance of the robust estimates are much less significant. iii) The differences in performance between direct and robust estimates are less significant (yet still clearly to the advantage of the latter group). These observations can be interpreted in two ways. First, they can be read as direct consequences of the restriction of the space of allowed solutions. Second, in consistency with results reported in [32], where it is concluded that constraints eases the optimization process, notably when S_n is ill-conditioned, short sale can be understood as an implicit *regularization* in estimation procedures. Indeed, in absence of short sale restrictions, Solution (3) to Problem (2) can be obtained via a classical Lagrangian formulation, $L(w,\lambda) = w'\Sigma w + \lambda\left(1_p' w - 1\right)$, (with λ a Lagrange multiplier), whose resolution is straightforward, yielding:

$$2\Sigma w + \lambda \cdot 1_p = 0.$$

In the presence of short sale constraints, the Lagrangian formulation becomes $L(w,\lambda,\nu) = w'\Sigma w + \lambda\left(1_p' w - 1\right) - \nu' w$, (with $\lambda \in \mathbb{R}$ and $\nu \in \mathbb{R}_+^p$ the Lagrange multipliers). Deriving the Karush-Kuhn-Tucker conditions of optimality yields (cf. Supplementary Material)

$$2\left[\Sigma + \mathrm{Diag}(\nu) - \frac{1}{2}\cdot\left(1_p\nu' + \nu 1_p'\right)\right] w + \lambda \cdot 1_p = 0\,, \quad (28)$$

which, compared to the unconstrained solution, reads the same but for the covariance matrix Σ being replaced by a *shrinked effective* surrogate, $\Sigma + \mathrm{Diag}(\nu) - \frac{1}{2}\cdot\left(1_p\nu' + \nu 1_p'\right)$. In nature, constraints act as a shrinkage leading to smaller effective correlations, thus limiting the need and impacts of further improvements in covariance or precision estimation.

5) $\hat{\Sigma}$ or $\hat{\Pi}$?: Factor and Shrinkage are the dominant classes for $\hat{\Sigma}$, while Sparsity and (to a lesser extent Factor) dominates for $\hat{\Pi}$. This suggests a generic behavior and robustness in the Factor approach.

Sparsity applied to Π appears as the estimation strategy performing best, irrespective of p/n and constraints. This is particularly clear for Dataset I, and still holds for Dataset II though Shrinkage on Σ and Sparsity on Π yield essentially equivalent performance. This can likely be well explained by the structure of (notably) Dataset I, which consists in shares from large capitalization European firms, that is in homogeneous assets, which, according to financial theory, are

TABLE IV
GMVP Performance – Dataset II: Fama-French Portfolios – Π. Same legend as table I.

	$n=30$, $p/n \simeq 3$				$n=90$, $p/n \simeq 1$			
	V	S	TO	H^{-1}	V	S	TO	H^{-1}
	Π - Long/Short							
	Factor							
$\hat{\Pi}^{(4)}$	9.41 (0.28)	2.16 (0.15)	1.22 (0.88)	5.13 (3.60)	9.10 (0.27)	2.19 (0.18)	0.40 (0.29)	5.62 (3.23)
$\hat{\Pi}^{(5)}$	41.02 (7.98)	0.58 (0.11)	0.70 (7.59)	64.99 (14.81)	8.55 (0.28)	2.34 (0.18)	0.41 (0.29)	5.43 (2.67)
	Shrinkage							
$\hat{\Pi}^{(7)}$	N/A	N/A	N/A	N/A	16.20 (0.26)	0.89 (0.13)	2.50 (2.83)	34.65 (26.26)
	Sparsity							
$\hat{\Pi}^{(9)}$	**6.86** (0.17)	**2.68** (0.18)	**1.29** (0.42)	**5.61** (2.35)	**6.79** (0.18)	**2.70** (0.17)	**0.57** (0.25)	**5.23** (1.83)
	Π - Long Only							
	Factor							
$\hat{\Pi}^{(4)}$	23.97 (0.36)	0.55 (0.13)	0.58 (0.30)	5.83 (4.80)	15.26 (0.23)	0.98 (0.13)	0.02 (0.01)	87.29 (2.27)
$\hat{\Pi}^{(5)}$	17.90 (0.36)	0.88 (0.13)	0.21 (0.31)	64.06 (14.81)	10.67 (0.19)	2.18 (0.15)	0.31 (0.41)	1.88 (1.23)
	Shrinkage							
$\hat{\Pi}^{(7)}$	N/A	N/A	N/A	N/A	15.22 (0.26)	0.99 (0.14)	0.84 (0.31)	57.52 (18.93)
	Sparsity							
$\hat{\Pi}^{(9)}$	**9.38** (0.13)	**2.11** (0.14)	**0.57** (0.28)	**6.13** (3.44)	**9.37** (0.12)	**2.13** (0.13)	**0.21** (0.14)	**6.49** (3.25)

sensitive to a small number of common economic factors. They hence show low partial correlation and thus a sparse precision matrix. Sparsity on Π precludes sparsity on Σ thus likely explaining the dramatically poor performance of sparsity on Σ. Further, this factorial structure is also likely explaining the satisfactory performance of factor based estimates for Σ.

IV. Conclusions and perspectives

We have conducted an in-depth study of the relative performance of different estimation strategies of the GMVP based on the inversion of the estimated covariance matrix or the direct estimation of the precision matrix.

All together, the results reported here tend to show that robust estimates outperform significantly direct estimates achieving improved trade-offs between a poor estimate of the full covariance structure (via the sample covariance estimator) and neglecting all dependencies (by keeping only its diagonal). Further, factor based estimates perform satisfactorily both for Σ and Π. Sparsity precision-based asset allocation yield on average the best performance across datasets, conditions and ratios p/n, and globally when gathering all indices. Notably for large p/n, sparsity precision-based asset allocations degrade in performance less than other strategies This may stem from the fact that short estimation windows avoid estimation blurring by data non-stationarity, while sparsity permits to maintain a sufficient quality for risk assessment. Robustness along time of both the procedures and conclusions reported here have been checked by splitting the 60-year long Dataset II into three sub-periods of 20 years, performance tables are reported in Supplementary Material, Section VI.

We think that these empirical results are of interest both from academic and professional points of view in so far as they pave the way toward the development of new estimation methods for optimal portfolio weights in the mean-variance framework and yield new questions regarding the informational content of the sample covariance and precision matrices. Notably, the latent factors approach, which is, to a large extend, related to the so-called random matrix theory, has recently been brought back to the front of the scene in [35]. It extends the factor approach to account for the fact that the actual number of factors is generally unknown. The idea consists in the identification of the significant eigenvalues and the eigenvectors of the covariance matrix, a notoriously difficult problem as soon as the ratio p/n of the number of assets to the number of observations is not small [50], [29], [19], [1]. Further, statistical principles can be combined to construct advanced classes of estimators, in the spirit of constructions studied in [23], [24], [54] combining e.g. Sparsity and Factor. These are under current investigation.

Appendix A
Shrinkage of the covariance matrix

In order to prove lemma 2, we follow [11]. We are looking for the parameters ρ and Δ, where Δ is a p-dimensional diagonal matrix, which minimize the quadratic loss function

$$L(\nu,\rho) = \mathrm{E}\left[||\rho\Delta + (1-\rho)S_n - \Sigma||^2\right] \tag{29}$$

where $||\cdot||$ denotes the Frobenius norm while S_n denotes the unbiased sample covariance matrix estimate from n iid random vectors of Gaussian assets returns with covariance matrix Σ.

The minimization of the quadratic loss function (29) with respect to Δ yields

$$\Delta = \mathrm{Diag}\,\Sigma, \tag{30}$$

i.e., Δ only retains the diagonal terms of the covariance matrix Σ, that will be estimated by

$$\hat{\Delta} = \mathrm{Diag}\,S_n. \tag{31}$$

After substitution in (29), the minimization with respect to ρ leads to

$$\rho = 1 - \frac{\mathrm{Tr}\left(\Sigma^2\right) - \mathrm{Tr}\left[(\mathrm{Diag}\,\Sigma)^2\right]}{\mathrm{E}\left[\mathrm{Tr}\left(S_n^2\right)\right] - \mathrm{E}\left[\mathrm{Tr}\left((\mathrm{Diag}\,S_n)^2\right)\right]}. \tag{32}$$

Notice that, up to now, this derivation is totally free from the distributional properties of the sample matrix S_n apart from the absence of bias.

Let us now use the fact that the sample covariance matrix follows a Wishart distribution $(n-1)\cdot S_n \sim \mathcal{W}_p(n-1,\Sigma)$, so that [42]

$$\mathrm{Cov}\left((S_n)_{ij},(S_n)_{kl}\right) = \frac{1}{n-1}\begin{pmatrix}\Sigma_{ik}\cdot\Sigma_{jl} \\ +\Sigma_{il}\cdot\Sigma_{jk}\end{pmatrix}. \tag{33}$$

As a consequence

$$\mathrm{E}\left[\mathrm{Tr}\left(S_n^2\right)\right] = \frac{n}{n-1}\mathrm{Tr}\left(\Sigma^2\right) + \frac{1}{n-1}\left(\mathrm{Tr}\,\Sigma\right)^2, \tag{34}$$

and

$$\mathrm{E}\left[\mathrm{Tr}\left((\mathrm{Diag}\,S_n)^2\right)\right] = \frac{n+1}{n-1}\mathrm{Tr}\left[(\mathrm{Diag}\,\Sigma)^2\right]. \tag{35}$$

Notice that these relations can straightforwardly be obtained by application of the Stein-Haff identity. By substitution of equations (34) and (35) in (32), we obtain the oracle shrinkage estimator

$$\rho = \frac{\mathrm{Tr}\left(\Sigma^2\right) + \left(\mathrm{Tr}\,\Sigma\right)^2 - 2\mathrm{Tr}\left[\left(\mathrm{Diag}\,\Sigma\right)^2\right]}{n\mathrm{Tr}\left(\Sigma^2\right) + \left(\mathrm{Tr}\,\Sigma\right)^2 - (n+1)\mathrm{Tr}\left[\left(\mathrm{Diag}\,\Sigma\right)^2\right]}. \quad (36)$$

In order to derive an estimator of ρ, we follow the line of [11] and introduce the Oracle Approximating Shrinkage (OAS) estimator as the limit of the iterative process

$$\hat{\Sigma}_j = \hat{\rho}_j \cdot \mathrm{Diag}\,S_n + (1-\hat{\rho}_j)\cdot S_n, \quad (37)$$

and

$$\hat{\rho}_{j+1} = \frac{\mathrm{Tr}\left(\hat{\Sigma}_j \cdot S_n\right) + \left(\mathrm{Tr}\,\hat{\Sigma}_j\right)^2 - 2\mathrm{Tr}\left[\left(\mathrm{Diag}\,\hat{\Sigma}_j\right)^2\right]}{n\mathrm{Tr}\left(\hat{\Sigma}_j \cdot S_n\right) + \left(\mathrm{Tr}\,\hat{\Sigma}_j\right)^2 - (n+1)\mathrm{Tr}\left[\left(\mathrm{Diag}\,\hat{\Sigma}_j\right)^2\right]}. \quad (38)$$

By substitution of (37) into (38) we get

$$\rho_{j+1} = \frac{1-\phi_n\cdot\rho_j}{1+(n-1)\cdot\phi_n - n\phi_n\cdot\rho_j}\,, \quad (39)$$

with

$$\phi_n = \frac{\mathrm{Tr}\left(S_n^2\right) - \mathrm{Tr}\left[\left(\mathrm{Diag}\,S_n\right)^2\right]}{\mathrm{Tr}\left(S_n^2\right) + \left(\mathrm{Tr}\,S_n\right)^2 - 2\mathrm{Tr}\left[\left(\mathrm{Diag}\,S_n\right)^2\right]}\,, \quad (40)$$

and $0 \leq \phi_n \leq 1$ by construction. Taking the limit as $j\to\infty$ we obtain the result stated in Lemma 2.

Appendix B
Shrinkage of the precision matrix

As for the shrinkage estimator of the precision matrix toward identity, the quadratic loss function becomes

$$L(\nu,\rho) = \mathrm{E}\left[||\rho\nu Id + (1-\rho)P_n - \Pi||^2\right], \quad (41)$$

where $\Pi = \Sigma^{-1}$ and P_n is the unbiased sample precision matrix obtained by inversion of the sample covariance matrix S_n if $n>p$ or is given by the Moore-Penrose generalized inverse if $n\leq p$. The minimization with respect to ν yields

$$\hat{\nu} = \frac{1}{p}\mathrm{Tr}\,P_n\,. \quad (42)$$

As a consequence, by substitution in (41) and minimization with respect to ρ we get

$$\rho = \frac{\mathrm{E}\left[\mathrm{Tr}\left(P_n^2\right)\right] - \mathrm{Tr}\left(\Pi^2\right) - \frac{1}{p}\mathrm{Var}\left(\mathrm{Tr}\,P_n\right)}{\mathrm{E}\left[\mathrm{Tr}\left(P_n^2\right)\right] - \frac{1}{p}\left(\mathrm{Tr}\,\Pi\right)^2 - \frac{1}{p}\mathrm{Var}\left(\mathrm{Tr}\,P_n\right)}. \quad (43)$$

In the case $n>p$, the inverse of the sample covariance matrix exists and the unbiased sample estimator of the precision matrix is (provided that $n>p+2$)

$$P_n = \frac{n-p-2}{n-1}\cdot S_n^{-1} \quad (44)$$

since $(n-p-2)^{-1}\cdot P_n = ((n-1)\cdot S_n)^{-1} \sim \mathcal{W}_p^{-1}(n-1,\Pi)$, where $\mathcal{W}^{-1}$ is the inverse Wishart distribution.

From [48], we know that

$$\mathrm{Cov}\left(\left(P_n\right)_{ij}, \left(P_n\right)_{kl}\right) = \frac{2\cdot\Pi_{ij}\cdot\Pi_{kl} + (n-p-2)\left(\Pi_{ik}\cdot\Pi_{jl} + \Pi_{il}\cdot\Pi_{jk}\right)}{(n-p-1)(n-p-4)}\,, \quad (45)$$

provided that $n>p+4$. Hence

$$\mathrm{E}\left[\mathrm{Tr}\left(P_n^2\right)\right] = \left(\frac{n-p}{(n-p-1)(n-p-4)} + 1\right)\mathrm{Tr}\left(\Pi^2\right) + \frac{n-p-2}{(n-p-1)(n-p-4)}\left(\mathrm{Tr}\,\Pi\right)^2, \quad (46)$$

and

$$\mathrm{Var}\left(\mathrm{Tr}\,P_n\right) = 2\cdot\frac{\left(\mathrm{Tr}\,\Pi\right)^2 + (n-p-2)\cdot\mathrm{Tr}\left(\Pi^2\right)}{(n-p-1)(n-p-4)}. \quad (47)$$

Thus, the oracle shrinkage parameter is

$$\rho = \frac{\frac{n-p-\frac{2}{p}(n-p-2)}{(n-p-1)(n-p-4)}\cdot\mathrm{Tr}\left(\Pi^2\right) + \frac{n-p-2-\frac{2}{p}}{(n-p-1)(n-p-4)}\cdot\left(\mathrm{Tr}\,\Pi\right)^2}{\left[1+\frac{n-p-\frac{2}{p}(n-p-2)}{(n-p-1)(n-p-4)}\right]\mathrm{Tr}\left(\Pi^2\right) + \left[\frac{n-p-2-\frac{2}{p}}{(n-p-1)(n-p-4)} - \frac{1}{p}\right]\left(\mathrm{Tr}\,\Pi\right)^2}. \quad (48)$$

As previously, we obtain the OAS estimator as the limit of the iterative process

$$\hat{\Pi}_j = \hat{\rho}_j\cdot\frac{1}{p}\mathrm{Tr}\,P_n\cdot Id + (1-\hat{\rho}_j)\cdot P_n, \quad (49)$$

and

$$\hat{\rho}_{j+1} = \frac{\left[n-p-\frac{2}{p}(n-p-2)\right]\mathrm{Tr}\left(\hat{\Pi}_j P_n\right) + \left(n-p-2-\frac{2}{p}\right)\left(\mathrm{Tr}\,\hat{\Pi}_j\right)^2}{\left[n-p-\frac{2}{p}(n-p-2) + (n-p-1)(n-p-4)\right]\mathrm{Tr}\left(\hat{\Pi}_j P_n\right) + \left[n-p-2-\frac{2}{p} - \frac{1}{p}(n-p-1)(n-p-4)\right]\left(\mathrm{Tr}\,\hat{\Pi}_j\right)^2}. \quad (50)$$

The limit as $j\to\infty$ provides the result stated in Lemma 3. The proof of lemma 4 follows exactly the same line; it is omitted.

References

[1] A.N. Akansu and M.U. Torun, (2015) A Primer for Financial Engineering: Financial Signal Processing and Electronic Trading. *Elsevier*.

[2] Bengtsson, C., and J. Holst (2003) On Portfolio Selection: Improved Covariance Matrix Estimation for Swedish Asset Returns. *Working paper* Lund University.

[3] Bien, J., and R. J. Tibshirani (2011) Sparse Estimation of a Covariance Matrix. *Biometrika* **98**, 807–820.

[4] Bickel, P. J., and E. Levina (2008) Covariance regularization by thresholding. *Annals of Statistics* **36(6)**, 2577–2604.

[5] Bodnar, T., and Y. Okhrin (2008) Properties of the Singular, Inverse and Generalized Inverse Partitioned Wishart Distributions. *Journal of Multivariate Analysis* **99**, 2389–2405.

[6] Boyd, S., N. Parikh, E. Chu, B. Peleato and J. Eckstein (2010) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning* **3**, 1-122.

[7] Cai, T., W. Liu, (2011) Adaptive Thresholding for Sparse Covariance Matrix Estimation. *Journal of the American Statistical Association* **106 (494)**, 672–684.

[8] Cai, T., W. Liu, and X. Luo (2011) A Constrained l_1 Minimization Approach to Sparse Precision Matrix Estimation. *Journal of the American Statistical Association* **106**, 594–607.

[9] Carhart, M. M. (1997) On Persistence in Mutual Fund Performance. *Journal of Finance* **52**, 57–82.

[10] Chamberlain, G., and M. Rothschild (1983) Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets. *Econometrica* **51**, 1305–1324.

[11] Chen, Y., A. Wiesel, Y. C. Eldar and A. O. Hero (2010) Shrinkage Algorithms for MMSE Covariance Estimation. *IEEE Transactions on Signal Processing* **58**, 5016–5029.

[12] Chen, Y., A. Wiesel and A. O. Hero (2011) Robust shrinkage estimation of high dimensional covariance matrices. *IEEE Transactions on Signal Processing* **59**, 4097–4107.

[13] Cook, R. D., and L. Forzani (2011) On the Mean and Variance of the Generalized Inverse of a Singular Wishart Matrix. *Electronic Journal of Statistics* **5**, 146–158.

[14] Dahl, J., L. Vandenberghe and V. Roychowdhury (2008) Covariance selection for non-chordal graphs via chordal embedding. *Optimization Methods and Software* **23**, 501–520.

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

[16] DeMiguel, V., Martin-Utrera, A., & Nogales, F. J. (2013) Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *ournal of Banking & Finance* **37(8)**, 3018-3034.

[17] Dempster, A. P. (1972) Covariance Selection. *Biometrics* **28**, 157–75

[18] Disatnik, D. J., and S. Benninga (2007) Shrinking the Covariance Matrix – Simpler is Better. *Journal of Portfolio Management* **33(4)**, 55–63.

[19] El Karoui, N. (2008) Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, **36**, 2757–2790.

[20] Elton, E. J., and M. J. Gruber (1973) Estimating the Dependence Structure of Share Prices – Implications for Portfolio Selection, *Journal of Finance* **28**, 1203–1232.

[21] Fabozzi, F. J. and H. M. Markowitz (2011) *The Theory and Practice of Investment Management: Asset Allocation, Valuation, Portfolio Construction, and Strategies* (Wiley, 2nd Edition).

[22] Fama, E. F., and K. R. French (1992) The Cross-Section of Expected Stock Returns. *Journal of Finance* **47**, 427–465.

[23] Fan, J. , Liao, Y., and Mincheva, M. (2011), High Dimensional Covariance Matrix Estimation in Approximate Factor Models, *The Annals of Statistics* **39**, 3320-3356.

[24] Fan, J., Liao, Y. and Mincheva, M. (2013), Large Covariance Estimation by Thresholding Principal Orthogonal Complements (with discussion), *Journal of Royal Statistical Society B* **75**, 603-680.

[25] Friedman, J., T. Hastie, and R. Tibshirani (2008) Sparse Inverse Covariance Estimation with the Graphical Lasso. *Biostatistics* **9**, 432–441.

[26] Hastie, T., J. Friedman, and R. Tibshirani (2009) *The elements of statistical learning*, (Springer, 2nd edition).

[27] Goto, S., & Xu, Y. (2016) Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis (JFQA)* **50**, 1415-1441.

[28] Haff, L. R. (1979) Estimation of the Inverse Covariance Matrix: Random Mixtures of the Inverse Wishart Matrix and the Indentity. *Annals of Statistics* **7**, 1264-1276.

[29] Haff, L. R. (1991) The Variational Form of Certain Bayes Estimators. *Annals of Statistics* **19**, 1163–1190.

[30] Hero, A. O., and B. Rajaratnam (2011) Large Scale Correlation Screening. *Journal of the American Statistical Association* **106**, 1540–1552.

[31] Hsieh, C.-J., M. A. Sustik, I. S. Dhillon and P. Ravikumar (2011) Sparse Inverse Covariance Matrix Estimation Using Quadratic Approximation. *Advances in Neural Information Processing Systems* **24**, 2330–2338.

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

[33] Kourtis, A., G. Dotsis, and R. N. Markellos (2012) Parameter Uncertainty in Portfolio Selection: Shrinking the Inverse Covariance Matrix. *Journal of Banking and Finance* **36**, 2522–2531.

[34] Kubokawa, T. and M. S. Srivastava (2008) Estimation of the Precision Matrix of a Singular Wishart Distribution and its Application in High-Dimensional Data. *Journal of Multivariate Analysis* **99**, 1906–1928.

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

[36] Ledoit, O., and M. Wolf (2004b) Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection. *Journal of Empirical Finance* **10**, 603–621.

[37] Ledoit, O., and M. Wolf (2012) Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices. *Annals of Statistics* **40**, 1024–1060.

[38] Lian, H. (2011) Shrinkage Tuning Parameter Selection in Precision Matrices Estimation. *Journal of Statistical Planning and Inference* **141**, 2839–2848.

[39] Marcenko, V. A., and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics* **1**, 457–483

[40] Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments* (John Wiley & Sons).

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

[42] Muirhead, R. J. (2005) *Aspects of Multivariate Statistical Theory* (Wiley-Interscience, 2nd edition).

[43] Ravikumar, P. , M. Wainwright, G. Raskutti and B. Yu (2011) High-dimensional co-variance estimation by minimizing l1-penalized log-determinant divergence, Electronic Journal of Statistics **5**, 935–980.

[44] Rothman, A. J. , E. Levina, and J. Zhu (2009) Generalized thresholding of large covariance matrices. *Biometrika* **104**, 177–186.

[45] Rothman, A. J. (2012) Positive definite estimators of large covariance matrices. *Biometrika* **99**, 733–740.

[46] Schäfer, J. and K. Strimmer (2005) A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology* **4**, DOI: 10.2202/1544-6115.1175.

[47] Sharpe, W. (1963) A Simplified Model for Portfolio Analysis. *Management Science* **79**, 277–231.

[48] Siskind, V. (1972) Second Moments of Inverse Wishart-Matrix Elements. *Biometrika* **59**, 690–691.

[49] Stein, C. (1956) Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **1**, 197–206.

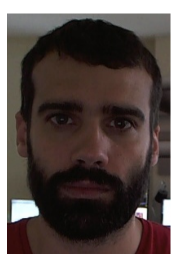
[50] Stein, C. (1975) Estimation of a Covariance Matrix. *Rietz Lecture, 39th Annual Meeting IMS*. Atlanta, Georgia.

[51] Tanaka, M., and K. Nakata (2013) Positive definite matrix approximation with condition number constraint. *Optimization Letters*, DOI:10.1007/s11590-013-0632-7.

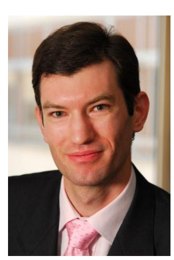
[52] Won, J. H., and S. J. Kim (2006) Maximum likelihood covariance estimation with a condition number constraint. *Fortieth Asilomar Conference on Signals, Systems, and Computers*, 1445–1449

[53] E. Yang, A. C. Lozano, and P. Ravikumar (2014) Elementary estimators for sparse covariance matrices and other structured moments. *In International Conference on Machine learning (ICML), 31, 2014*

[54] L. Yang, R. Couillet, M. McKay (2015), A Robust Statistics Approach to Minimum Variance Portfolio Optimization, IEEE Transactions on Signal Processing, **63(24)**, 6684–6697.



Marc Senneret is Head of Applied Research at Vivienne Investissement, Lyon. He holds a PhD in Physics from Université Diderot, Paris.



Yannick Malevergne is Professor of finance at Université Jean Monnet, Saint-Etienne. He was previously Senior Researcher with the Chair of Entrepreneurial Risks at ETH Zurich. He holds a Ph.D in Physics from the University of Nice. His core research is devoted to the study of extreme risks in financial markets, and to the development of new portfolio management methods and asset pricing models. He has authored or co-authored many articles both in academic reviews and professional journals as well as a books on extreme risks.

Patrice Abry is currently CNRS Senior Scientist, at the Physics dept. of ENS de Lyon. He also is Professeur-Agrégé in Physical Sciences Physiques since 1989 and hold a PhD in Physics and Signal Processing from Université Claude-Bernard University, Lyon for which he received the AFCET-MESR-CNRS prize. Patrice Abry has been elected IEEE Fellow in 2011. His research interests include multiscale and multifractal, multivariate and multidimensional analysis and representations, with applications ranging from Internet traffic to heart rate variability.

Gérald Perrin holds a PhD in fluid mechanics from the Ecole Centrale de Lyon, worked for many years in private research centers where he developed advanced skills in the modeling and simulation of complex processes. With this dual expertise in programming and nonlinear systems analysis, he joined Vivienne Investissement where he built a secure IT infrastructure.

Laurent Jaffrès holds an MBA from EMLyon. His current research interests include Quantitative Finance and Robust Portfolio Construction. In 2005, he founded Vivienne Investissement, a quantitative asset-management company, of which he is the Chairman and CEO. Vivienne Investissement's goal is to provide customers with the best of research, aiming at innovation to bring strength and to constantly improve risk understanding and management processes. The main idea is to make the data speak in the context of understanding the facts.