

Covariance versus Precision Matrix Estimation for Efficient Asset Allocation

M. Senneret*, Y. Malevergne†, P. Abry‡, G. Perrin* and L. Jaffrès*

* Vivienne Investissement, Lyon, France

† Coactis EA 4161 – Université de Lyon, and EMLYON Business School, France

‡ CNRS, ENS Lyon, Physics Dept., France

Abstract—Asset allocation constitutes one of the most crucial and most challenging task in financial engineering. In many allocation strategies, the estimation of large covariance or precision matrices from short time span multivariate observations is a mandatory yet difficult step. In the present contribution, a large selection of elementary to advanced estimation procedures for the covariance as well as for precision matrices, are organized into classes of estimation principles, reviewed and compared. To complement this overview, several additional estimators are explicitly derived and studied theoretically. Rather than estimation performance evaluated from synthetic simulated data, performance of the estimation procedures are assessed empirically by financial criteria (volatility, Sharpe ratio,...) quantifying the quality of asset allocation in the mean-variance framework. Performance are quantified by application to a large set of about 250 European stock returns across the last 15 years, up to August 2015. Several real-life scenarios are considered, with different ratios of the estimation time span to the size of the basket of candidate assets. This large scale comparative study allows us to address issues such as the relative benefits and difficulties of using direct estimates of the precision matrix over the estimation of the covariance matrix. Also, we analyze the impacts of realistic constraints, such as short sale restrictions.

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], [6], [20]. Amongst the numerous available strategies, the mean-variance framework of Markowitz [37], [38] 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 $\{r_k(t), k = 1, \dots, p, t = 1, \dots, n\}$, the GMVP approach amounts to estimating the vector of weights $\{w_k, k = 1, \dots, p\}$, that minimizes the overall standard deviation (hereafter referred to as the *volatility*) of the portfolio return:¹

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

Without restriction on short-sales, i.e., weights w_k can be either positive or negative, the GMVP is obtained as the

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

solution to the following optimization program:

$$w^* = \underset{w}{\operatorname{argmin}} w' \Sigma w, \text{ s.t. } 1_p' w = 1, \text{ with } 1_p' = \underbrace{[1, \dots, 1]}_{p\text{-times}}, \quad (2)$$

where $'$ 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]):

$$w^* = (1_p' \Pi 1_p)^{-1} \Pi 1_p. \quad (3)$$

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 [15], [19].

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, variance in estimation implies significant uncertainties on eigenvalues and eigenvectors; 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 [39], to the point that they may underperform equally weighted portfolio (EWP) strategies, that simply ignore covariance estimation and use only a diagonal identity matrix instead (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*².

Related work. A large amount of research efforts has been dedicated to overcoming such limitations. 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 S_n [32], while the details of the shrinkage themselves have little impact [17]. However, despite reduced volatility, the benefits of

²In the finance literature and econometrics, *out-of-sample* refers to performance computed on data that were not used for parameter estimation, tuning or learning.

shrinkage based estimates in terms of performance measured by financial criteria such as the Sharpe ratio (i.e., the ratio of Gain to the volatility) are still questioned (cf. e.g., [27]).

Along another line, one can surprisingly note that, in the last decade, most research efforts (but for the noticeable exception of [28]) were dedicated to improve the estimation of Σ , while the key quantity is obviously Π . It is however well-known that the inverse of S_n yields a poor estimate of the precision matrix, notably when $n \simeq p$ [40].

Further, restrictions on short-sales which amount to account for the additional constraint $w_k \geq 0, \forall k$ in problem (2) 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 [27]. The impact of such additional constraints remains yet undocumented when a direct estimate of Π is used.

Goals, contributions and outline. In this context, the overall goals of the present contribution are to contribute to a large size study of covariance/precision matrix estimation strategies in asset allocation and to compare their performances. Notably, it intends to quantify which strategy yields optimal GMVP performance, as measured with financial metrics and in realistic contexts (e.g., when $n \simeq p$). More precisely the following issues will be addressed: i) Should one invert a covariance estimate or prefer a direct estimate of the precision ? ii) Which estimate of Σ , of Π should be preferred ? in what context ? iii) How do short sale restrictions or other additional financial constraints impact performances ? To achieve these goals, the first originalities of the present contribution are both to review and organize existing estimators into classes of estimation principles (direct, factors, shrinkage and sparsity) as well as to derive closed-form expressions for several reliable estimators for Σ and Π that had never been made explicit so far (cf. Section II). A 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 a very large data set of actual risky assets, which constitute a relevant basket that can be used in practice by investment firms: The $p = 244$ largest capitalizations of the STOXX Europe 600 index, across more than 15 years, up to August 2015. Several realistic situations will be explored: $n \gg p$, $n \geq p$, $n \gtrsim p$ and $n \leq p$. Setting and results are reported and commented in Section III, together with the impact of short sale restrictions on not only the estimates of Σ but also on those of Π . Future works and conclusions are envisaged in Section IV.

II. COVARIANCE AND PRECISION ESTIMATION STRATEGIES

This section presents the estimators for Σ and Π retained in this study, organized into four classes (direct, factor based,

shrinkage based, sparsity based), and provides an overview of their main properties. It also details original results for the estimation of both the covariance and precision matrices.

A. Direct sample estimates

The sample covariance matrix estimator S_n (reabeled here $\hat{\Sigma}^{(1)}$ for consistency):

$$\hat{\Sigma}^{(1)} := \frac{1}{n-1} \left[\sum_{t=1}^n r_t \cdot r_t' - \frac{1}{n} \left(\sum_{t=1}^n r_t \right) \cdot \left(\sum_{t=1}^n r_t' \right) \right], \quad (4)$$

where r_t denotes the p -vector of asset returns on day t , is certainly the simplest estimator one can consider. However, it suffers from two main deficiencies. First, when the number of observations n is less than the number of assets p , the sample covariance matrix is not full rank, hence it is not invertible. In such a case, the Moore-Penrose generalized inverse is usually retained to estimate the precision matrix

$$\hat{\Pi}^{(1)} := \left[\hat{\Sigma}^{(1)} \right]^+, \quad (5)$$

where \cdot^+ denotes the generalized inverse [2]. Second, even if the sample covariance matrix is full rank, its inverse only provides a biased estimator of the inverse population covariance matrix.

Many simple, and sometimes naive but efficient, alternatives have been proposed. Among many others, let us refer to the replacement of the sample covariance matrix by a scalar – and actually the identity – matrix

$$\hat{\Sigma}^{(2)} := \text{Id}_p \quad \text{and} \quad \hat{\Pi}^{(2)} := \text{Id}_p, \quad (6)$$

in which case GMVP simplifies to EWP, or by the diagonal matrix of the sample variances

$$\hat{\Sigma}^{(3)} := \text{diag} \left(\hat{\Sigma}^{(1)} \right) \quad \text{and} \quad \hat{\Pi}^{(3)} := \left[\hat{\Sigma}^{(3)} \right]^{-1}. \quad (7)$$

Let us remark that 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 sample covariance matrix and its (generalized) inverse will provide the benchmark strategy for the horse race in section III.

B. Factor estimates

1) *Factor principle:* Alternatively, in order (i) to reduce the noise in the sample covariance matrix, (ii) to get a full rank estimate of the covariance matrix even when the number of assets is larger than the number of observations and (iii) to get a reliable estimate of the inverse covariance matrix, the factor models can provide a versatile approach. The simplest case is derived from the Sharpe market model [44] in which the return on the market portfolio is assumed to be the single relevant factor. More general models based on three or four factors can provide better approximations to the actual covariance matrix for stock returns [9], [21]. Alternatively, when the factors are unknown or unobservable, an approximate factor structure can be reconstructed from a principal component analysis or a singular value analysis [10], [35].

2) *Exogenous factors*: For simplicity we will consider the single index market model as representative of this first class of models:

$$r_k(t) = \alpha_k + \beta_k \cdot r_m(t) + \varepsilon_k(t), \quad (8)$$

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 in the present study, on day t . The vector parameter $\alpha = \{\dots, \alpha_k, \dots\}$ and $\beta = \{\dots, \beta_k, \dots\}$ are the p -vectors of intercepts and factor loadings while $\varepsilon(t) = \{\dots, \varepsilon_k(t), \dots\}$ is the p -vector of idiosyncratic residuals with diagonal covariance matrix Δ . For simplicity, let us assume iid normal observations. The covariance and precision matrices read

$$\Sigma = \Delta + \sigma_m^2 \beta \beta' \quad \text{and} \quad \Pi = \Delta^{-1} - \frac{\Delta^{-1} \beta \cdot \beta' \Delta^{-1}}{\frac{1}{\sigma_m^2} + \beta' \Delta^{-1} \beta}, \quad (9)$$

where $\sigma_m^2 = \text{Var } r_{m,t}$.

One can obtain *plug-in* estimators by naive replacement of the parameters β and Δ by their Ordinary Least Square (OLS) unbiased estimators. However, these plug-in estimators do not result in unbiased estimators for Σ and Π .

Accounting for 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 get

$$\text{E} \left[\hat{\sigma}_m^2 \hat{\beta} \hat{\beta}' \right] = \sigma_m^2 \beta \beta' + \frac{1}{n-1} \cdot \Delta, \quad (10)$$

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:

$$\hat{\Sigma}^{(4)} := \hat{\sigma}_m^2 \hat{\beta} \hat{\beta}' + \frac{n-2}{n-1} \cdot \hat{\Delta}. \quad (11)$$

We can note that $\hat{\Sigma}^{(4)}$ only differs from the plug-in estimator by the multiplicative factor $(n-2)/(n-1)$ which derives from the fact that the entries of $\hat{\Delta}$ are χ^2 -distributed with $n-2$ degrees of freedom. Note also that, 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 (11).

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

$$(1 + \sigma_m^2 \cdot \beta' \Delta^{-1} \beta) \cdot \Delta^{-1} - \sigma_m^2 \cdot \Delta^{-1} \beta \cdot \beta' \Delta^{-1}. \quad (12)$$

As a consequence of the independence of the OLS estimators of β and Δ under normality, we get

$$\text{E} \left[\hat{\sigma}_m^2 \left(\hat{\beta}' \hat{\Delta}^{-1} \hat{\beta} \right) \hat{\Delta}^{-1} - \hat{\sigma}_m^2 \hat{\Delta}^{-1} \hat{\beta} \hat{\beta}' \hat{\Delta}^{-1} \right] = \left(\frac{n-2}{n-4} \right)^2 \cdot \left(\sigma_m^2 \left(\beta' \Delta^{-1} \beta \right) \Delta^{-1} - \sigma_m^2 \Delta^{-1} \beta \beta' \Delta^{-1} + \frac{p-1}{n-1} \Delta^{-1} \right) \quad (13)$$

while

$$\text{E} \left[\hat{\sigma}_m^2 \cdot \hat{\beta}' \hat{\Delta}^{-1} \hat{\beta} \right] = \frac{n-2}{n-4} \left[\sigma_m^2 \cdot \beta' \Delta^{-1} \beta + \frac{p}{n-1} \right]. \quad (14)$$

Hence a suitable estimator of the precision matrix reads:

$$\hat{\Pi}^{(4)} := \frac{n-4}{n-2} \cdot \hat{\Delta}^{-1} - \frac{\left(\frac{n-4}{n-2} \right)^2 \hat{\sigma}_m^2 \hat{\Delta}^{-1} \hat{\beta} \hat{\beta}' \hat{\Delta}^{-1} - \frac{\hat{\Delta}^{-1}}{n-1}}{1 + \frac{n-4}{n-2} \cdot \hat{\sigma}_m^2 \hat{\beta}' \hat{\Delta}^{-1} \hat{\beta} - \frac{p}{n-1}}. \quad (15)$$

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

3) *Endogenous factors*: When the factors are unobservable, one has to rely on the principal component analysis (PCA) or the singular value decomposition (SVD) from which many different estimators can be derived. We define two couples of estimators.

For consistency with the previous approach, we consider $(\hat{\Sigma}^{(5)}, \hat{\Pi}^{(5)})$ whose definition follows that of $(\hat{\Sigma}^{(4)}, \hat{\Pi}^{(4)})$ 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, given that we are not able to derive theoretically an improved estimator for the precision matrix, involving more than one factor, $\hat{\Pi}^{(6)}$ will not be considered.

C. The shrinkage approach

1) *Shrinkage principle*: The shrinkage of the sample covariance matrix towards a target user supplied matrix M , amounts to replace S_n with the following linear combination:

$$\hat{\Sigma}^{shrink} = (1 - \rho) S_n + \rho M, \quad (16)$$

where ρ results from the minimization of the quadratic loss function, which achieves the best trade-off between the bias and the variance of the resulting estimator:

$$L(\rho, M) = \text{E} \left[\|\hat{\Sigma}^{shrink} - \Sigma\|^2 \right]. \quad (17)$$

This approach was originally introduced in [46] and provides an optimal mix between the sample estimate of the covariance/precision matrix and a target matrix. More recently, [32] considered shrinkage toward a scalar matrix and toward the covariance matrix implied by Sharpe's market model while [3] considered shrinking the sample covariance matrix toward the covariance matrix derived from a latent factor model estimated by principal component analysis³. All in all, [17] suggests that the simplest approach to shrinkage provides the best results. However the recent advances proposed in [11], [12] show that better approximations of the covariance matrix can be obtained on the basis of improved shrinkage parameters in particular in the case where the input data are fat-tailed. Alternatively, non-linear shrinkage methods either based on the introduction of an upper limit for the condition number of the estimated covariance matrix [48], [51] or on the Marcenko-Pastur equation [36] seem to provide significant improvements [31], [33].

³The shrinkage parameter for several classical models can be found in [43].

2) *Covariance matrix*: The case of linear shrinkage is considered here only. We use the Oracle Approximating Shrinkage (OAS) estimator introduced in [11] for the shrinkage toward the identity matrix, whose performance are actually close to those of the oracle shrinkage estimator whose implementation requires 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 unbiased sample covariance matrix estimator S_n , the Oracle Approximating Shrinkage estimator of the covariance matrix toward the identify matrix is⁴*

$$\hat{\Sigma}^{(7)} = \hat{\rho} \cdot \frac{\text{Tr } S_n}{p} \cdot \text{Id}_p + (1 - \hat{\rho}) \cdot S_n, \quad (18a)$$

$$\hat{\rho} = \min \left\{ \frac{\left(1 - \frac{2}{p}\right) \text{Tr} (S_n^2) + (\text{Tr } S_n)^2}{\left(n - \frac{2}{p}\right) \cdot \left[\text{Tr} (S_n^2) - \frac{(\text{Tr } S_n)^2}{p}\right]}, 1 \right\}. \quad (18b)$$

When $n \rightarrow \infty$, $\hat{\rho} \rightarrow 0$ showing that S_n does not need improvement by shrinkage. Conversely, for small n , the leftmost term within the curl-brackets in (18b) can be larger than one, indicating that S_n is so noisy that it cannot be reliably used.

This result can be generalized to the shrinkage of the covariance matrix toward a diagonal matrix, which, to the best of our knowledge, is a new result:

Lemma 2. *Under the assumption of lemma 1, the Oracle Approximating Shrinkage estimator of the covariance matrix toward a diagonal matrix is*

$$\hat{\Sigma}^{(8)} = \hat{\rho} \cdot \text{Diag } S_n + (1 - \hat{\rho}) \cdot S_n, \quad (19a)$$

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

Proof is detailed in Appendix A,

3) *Precision matrix*: The shrinkage approach can also be successfully applied to the estimation of the precision matrix, which may be more relevant than the application of the shrinkage to the covariance matrix itself since the solution to the mean-variance optimization program directly involves this former one. When the sample covariance matrix is well-conditioned, namely when the number of observations n is larger than the number of assets p , [23] provides several random shrinkage estimators that outperform the naive estimator obtained by inversion of the sample covariance matrix. The proposed strategy is, in essence, quite close to the strategy applied in [32] for the shrinkage of the covariance matrix. Now, when the sample covariance matrix is singular, so that the previous method does not apply, [29] recently provides a shrinkage method to improve on the classical Moore-Penrose generalized inverse.

⁴This statement is the same as the one given in [11] with the replacement $n \rightarrow n - 1$ to account for the fact the mean value is unknown in the present case and for a recurrent typo in section III.C of [11] in which one has to account for the replacement $\frac{n+1-2}{p} \rightarrow n + 1 - \frac{2}{p}$ and $\frac{1-2}{p} \rightarrow 1 - \frac{2}{p}$.

In the context of portfolio optimization, the shrinkage of the precision matrix has only been recently considered in [28] which proposes a non-parametric cross-validation method for the estimation of the shrinkage parameter. We depart from this approach and propose two closed-form OAS estimator for the precision matrix when the sample covariance matrix 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 Oracle Approximating Shrinkage estimator of the precision matrix toward the identify matrix is*

$$\hat{\Pi}^{(7)} = \hat{\rho} \cdot \frac{\text{Tr } P_n}{p} \cdot \text{Id} + (1 - \hat{\rho}) \cdot P_n, \quad (20a)$$

$$\hat{\rho} = \min \left\{ \frac{\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1} \cdot \text{Tr} (P_n^2) + \frac{n-p-2-\frac{2}{p}}{n-p-1} \cdot (\text{Tr } P_n)^2}{\left[\frac{n-p-\frac{2}{p}(n-p-2)}{n-p-1} + n - p - 4\right] \times \left[\text{Tr} (P_n^2) - \frac{(\text{Tr } P_n)^2}{p}\right]}, 1 \right\}. \quad (20b)$$

Lemma 4. *The Oracle Approximating Shrinkage estimator of the precision matrix Π toward a diagonal matrix when $n > p + 4$ is*

$$\hat{\Pi}^{(8)} = \hat{\rho} \cdot \text{Diag} (P_n) + (1 - \hat{\rho}) \cdot P_n, \quad (21a)$$

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

The proof for these two lemma are postponed in Appendix B.

The singular case is much more tedious to handle since the sample covariance matrix does not admit a regular inverse. The sample precision matrix is then usually estimated by help of the Moore-Penrose generalized inverse whose statistic follows the generalized inverse Wishart distribution [5]. Unfortunately, to the best of our knowledge, the moments of this distribution do not admit known closed-form expressions apart from the case of cross-sectionally uncorrelated returns [13]. Hence the derivation of the shrinkage estimators $\hat{\Pi}^{(7)}$ and $\hat{\Pi}^{(8)}$ for the precision matrix in the singular case is left to future works.

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 [7], [8], [14], [16], [22], [34]. Such a calling upon parsimony in that context may be motivated from two different origins: Either from an *a priori* sparse modeling choice or from estimation issues.

Assuming *a priori* a sparse dependence model, i.e., the fact that, beyond the diagonal terms, only a small (compared to $p(p-1)$) number of entries of covariance or precision matrices 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 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 *a priori* imposed to covariance or precision.

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 [25]: For large matrices, estimated from short sample size, i.e., when $n \gtrsim p$ 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.

In both cases – sparse modeling or estimation issues – the practical challenge is to decide how many and which non diagonal entries should be set to zero. There have been on-going efforts to address sparse matrix estimation issues, concentrating first on the precision matrix [7], [8], [14], [16], [22], [34] and more recently on the covariance matrix [4], [42].

2) *Precision matrix*: In essence, the estimation of sparse precision matrices relies on minimizing a cost function, consisting of a balance between a data fidelity term associated to the precision matrix and a penalty term aiming at promoting sparsity. A state-of-the-art formulation of this problem is now referred to as the *Graphical Lasso* [22]. It balances the negative log-likelihood function, thus relying on the Graphical Gaussian Model framework, and hence following the original formulation due to [16], with an l_1 penalization of the esti-

mated precision matrix:

$$\hat{\Pi}^{(9)} := \underset{\Pi}{\operatorname{argmin}} \operatorname{Tr}(S_n \Pi) - \log \det \Pi + \lambda \cdot \|\Pi\|_1, \quad (22)$$

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. (22) 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* [4] or *Alternating Direction Method of Multipliers* algorithms [7]. In the present contribution, use is made of this latter algorithm.

3) *Covariance matrix*: Sparsity can be imposed onto the covariance matrix through the same formulation:

$$\hat{\Sigma}^{(9)} := \underset{\Sigma}{\operatorname{argmin}} \operatorname{Tr}(S_n \Sigma^{-1}) + \log \det \Sigma + \lambda \cdot \|\Sigma\|_1, \quad (23)$$

which however consists of a non-convex problem and is hence far more difficult to solve. It has however been observed that the argument in Eq. (23) can actually be split into a concave and a convex function, and that minimization can thus be performed by a majorization-minimization algorithm [4].

III. EMPIRICAL RESULTS

A. Data set and performance assessment

1) *Data Set*: To review and compare performance in GMVP strategies, the 9 different estimators for Σ and Π , described and studied in Section II above, have been applied to the daily returns of the $p = 244$ largest capitalizations of the Euro STOXX600 index, for a period of 15 years, from 2000, May 1st to 2015, August 31st., i.e., for $n = 4000$ trading days, thus constituting a realistic and remarkable set of $244 \times 4000 = 976000$ observations.

2) *Strategy set up*: It is quite well-known that a crucial issue in portfolio allocation lies in stock market time series being highly non-stationary. This naturally raises the question of assessing the typical *stationarity time scale* that can be associated to a data set. 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 of the past n days. The non-stationarity issue thus translates into selecting the size n of the estimation sliding window to achieve optimal performance. As such 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. These window sizes are $n = 500 \gg p$ and $n = 375 \geq p$, where S_n is a full rank matrix, and two sizes that are more difficult in terms

of estimation, yet more realistic in terms of real-life portfolio allocations, $n = 250 \gtrsim p$, and even $n = 125 \leq p$, where S_n is highly singular. These choices correspond to 24, 18, 12 and 6 months of trading days respectively. Portfolios are rebalanced once a week. Back-tests start after 500 days for all window sizes n to ensure that they all have the same duration: $L = 3500$ days. 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 , that balances sparsity versus data fidelity. There has been several interesting work investigating automated and adaptive tunings of λ , e.g., [41]. 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 . Results are reported for λ yielding the best GMVP performance only, thus slightly favoring that class.

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 generalized inequalities of the form $w_k \geq 0$.

$$w^* = \underset{w}{\operatorname{argmin}} w' \Sigma w, \text{ s.t. } 1_p' w = 1, \text{ and } w \geq 0, \quad (24)$$

read as a generalized inequality over the nonnegative orthant.

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. 22 has been implemented after the procedure described in [26]. We obtained the sparse covariance estimator $\hat{\Sigma}^{(9)}$ in Eq. 23 from [4], 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 is known and thus that estimation is applied to synthetic data, following an a priori prescribed model. 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), defined as,

$$\text{TO} = \frac{1}{n-1} \sum_{t=1}^{n-1} \sum_{k=1}^p |w_k(t+1) - w_k(t)|, \quad (25)$$

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}), defined as,

$$H^{-1} = \left(\frac{1}{n} \sum_{t=1}^n \sum_{k=1}^p w_k^2(t) \right)^{-1}, \quad (26)$$

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

Performance are reported in Tables I ($n = 125$), II ($n = 250$), III ($n = 375$) and IV ($n = 500$). In these tables, the tag *N/A* is used to indicate when Precision estimators are either meaningless or not available: This is the case for the direct estimators for the Precision matrix that are identical to the inverse of the estimators for the Covariance. This is also the case, as explained in Section II-B, for the 2-factor Precision estimator, theoretically not available. This is finally the case in Table I, where $n \leq p$, for $\hat{\Pi}^{(7)}$ and $\hat{\Pi}^{(8)}$ as their moments do not possess known closed form expressions (cf. Section II-C3 and Appendix B). Every second line complements the empirical mean obtained by average over the $L = 3500$ for each given financial metrics by its empirical standard deviation.

1) *Within $\hat{\Sigma}$* : Tables I to IV show that performance, irrespective of the considered metrics, remain quite close for estimators within a same class. They also show that shrinkage-based and factor-based estimators outperform Direct estimators (but $\hat{\Sigma}^{(1)}$) as well as Sparsity-based estimators both in terms of Sharpe ratios and volatilities by a factor of about 2, for all n . $\hat{\Sigma}^{(1)}$ is dominated in terms of volatility when $n \leq p$ or $n \gtrsim p$, but exhibits comparable performance for large n , $n \geq p$ and $n \gg p$. It remains nevertheless dominated in terms of Sharpe ratio. For $n \leq p$, its very large standard deviation shows that $\hat{\Sigma}^{(1)}$ is not a robust and reliable estimator, despite its being computed from the generalized inverse. It is also worth noting that sparsity based estimators for Σ globally perform poorly, and dramatically so when $n \leq p$.

The dominating 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 yields more concentrated portfolios with higher turn-overs, compared to the portfolios obtained from factor estimates. For both classes of estimators, while diversification (H^{-1}) does not depend on the estimation window size n , but the turn-over decreases notably when $n \geq p$.

All together, the factor based estimators (notably the two factor estimator $\hat{\Sigma}^{(6)}$) yield the most satisfactory asset al-

TABLE I
GLOBAL MINIMUM VARIANCE PORTFOLIO – 125 DAYS

125 days - Long/Short								
Σ				Π				
	V	S	TO	H^{-1}	V	S	TO	H^{-1}
Direct								
1	38.24 (49.07)	0.51 (0.79)	9.00 (1508)	2.84 (0.17)	N/A	N/A	N/A	N/A
2	19.65 (0.49)	0.67 (0.28)	0.00 (0.00)	244.00 (0.00)	N/A	N/A	N/A	N/A
3	15.71 (0.42)	0.67 (0.25)	0.04 (0.02)	164.60 (21.59)	N/A	N/A	N/A	N/A
Factor								
4	9.31 (0.26)	1.19 (0.28)	0.19 (0.12)	29.78 (15.71)	9.71 (0.28)	1.15 (0.28)	0.22 (0.14)	26.44 (15.22)
5	9.42 (0.26)	1.09 (0.27)	0.18 (0.11)	34.28 (18.57)	85.77 (17.97)	0.44 (0.37)	3.07 (16.80)	64.57 (66.63)
6	8.56 (0.25)	1.11 (0.25)	0.25 (0.16)	29.09 (16.59)	N/A	N/A	N/A	N/A
Shrinkage								
7	8.07 (0.18)	1.18 (0.26)	0.98 (0.38)	15.26 (8.93)	N/A	N/A	N/A	N/A
8	7.43 (0.17)	1.14 (0.25)	0.97 (0.38)	10.23 (5.72)	N/A	N/A	N/A	N/A
Sparsity								
9	249.70 (31.50)	0.20 (0.70)	134.46 (98.48)	0.05 (0.08)	7.05 (0.19)	1.45 (0.33)	0.41 (0.16)	18.65 (8.75)
125 days - Long Only								
Direct								
1	28.03 (0.20)	0.35 (0.25)	0.49 (0.14)	96.40 (5.83)	N/A	N/A	N/A	N/A
2	19.65 (0.49)	0.67 (0.26)	0.00 (0.00)	244.00 (0.00)	N/A	N/A	N/A	N/A
3	15.74 (0.36)	0.67 (0.28)	0.04 (0.02)	165.57 (21.48)	N/A	N/A	N/A	N/A
Factor								
4	9.18 (0.25)	1.10 (0.25)	0.17 (0.11)	15.17 (11.48)	45.01 (1.18)	0.33 (0.26)	0.22 (0.16)	12.96 (37.55)
5	9.33 (0.25)	1.02 (0.27)	0.17 (0.10)	17.16 (14.59)	35.30 (0.95)	0.29 (0.29)	0.12 (0.10)	80.67 (53.95)
6	9.18 (0.19)	1.04 (0.26)	0.18 (0.12)	15.53 (13.25)	N/A	N/A	N/A	N/A
Shrinkage								
7	9.21 (0.22)	1.18 (0.25)	0.22 (0.12)	16.54 (10.62)	N/A	N/A	N/A	N/A
8	9.10 (0.22)	1.08 (0.27)	0.23 (0.13)	11.50 (7.13)	N/A	N/A	N/A	N/A
Sparsity								
9	12.41 (0.28)	0.85 (0.26)	0.30 (0.06)	68.88 (20.15)	8.96 (0.24)	1.16 (0.28)	0.19 (0.11)	13.88 (8.67)

TABLE II
GLOBAL MINIMUM VARIANCE PORTFOLIO – 250 DAYS

250 days - Long/Short								
Σ				Π				
	V	S	TO	H^{-1}	V	S	TO	H^{-1}
Direct								
1	82.35 (8.75)	-0.02 (0.40)	48.42 (33.90)	0.15 (0.13)	N/A	N/A	N/A	N/A
2	19.65 (0.41)	0.67 (0.28)	0.00 (0.00)	244.00 (0.00)	N/A	N/A	N/A	N/A
3	16.15 (0.37)	0.67 (0.27)	0.02 (0.01)	172.82 (17.92)	N/A	N/A	N/A	N/A
Factor								
4	10.02 (0.27)	1.18 (0.28)	0.11 (0.07)	28.24 (12.70)	10.30 (0.25)	1.16 (0.26)	0.11 (0.08)	26.47 (12.52)
5	10.00 (0.31)	1.11 (0.24)	0.10 (0.07)	32.59 (15.06)	10.09 (0.28)	1.10 (0.24)	0.10 (0.07)	33.00 (14.59)
6	9.05 (0.23)	1.22 (0.26)	0.15 (0.12)	26.20 (10.77)	N/A	N/A	N/A	N/A
Shrinkage								
7	8.63 (0.20)	0.91 (0.31)	1.12 (0.47)	8.12 (4.32)	19.62 (0.51)	0.62 (0.30)	0.26 (0.40)	223.66 (45.78)
8	8.15 (0.21)	0.85 (0.26)	1.06 (0.43)	6.64 (3.28)	18.39 (0.51)	0.56 (0.26)	0.74 (0.30)	87.94 (25.69)
Sparsity								
9	16.15 (0.38)	0.67 (0.23)	0.02 (0.01)	172.82 (17.92)	7.16 (0.16)	1.57 (0.26)	0.22 (0.11)	19.36 (7.96)
250 days - Long Only								
Direct								
1	9.44 (0.20)	1.13 (0.25)	0.21 (0.09)	15.10 (5.02)	N/A	N/A	N/A	N/A
2	19.65 (0.41)	0.67 (0.28)	0.00 (0.00)	244.00 (0.00)	N/A	N/A	N/A	N/A
3	16.18 (0.40)	0.67 (0.25)	0.02 (0.01)	173.70 (17.87)	N/A	N/A	N/A	N/A
Factor								
4	9.25 (0.25)	1.21 (0.25)	0.10 (0.07)	14.09 (8.09)	15.14 (0.67)	0.86 (0.26)	0.14 (0.17)	2.96 (3.18)
5	9.46 (0.25)	1.14 (0.26)	0.10 (0.06)	15.74 (10.07)	9.94 (0.22)	1.07 (0.28)	0.10 (0.07)	14.85 (11.78)
6	9.23 (0.26)	1.18 (0.23)	0.10 (0.08)	13.87 (7.51)	N/A	N/A	N/A	N/A
Shrinkage								
7	9.21 (0.23)	1.20 (0.28)	0.13 (0.08)	13.08 (6.32)	19.66 (0.51)	0.61 (0.26)	0.09 (0.11)	239.25 (13.10)
8	9.18 (0.22)	1.15 (0.23)	0.13 (0.08)	11.08 (5.42)	18.51 (0.46)	0.55 (0.24)	0.60 (0.14)	104.46 (24.00)
Sparsity								
9	16.18 (0.41)	0.67 (0.23)	0.02 (0.01)	173.70 (17.87)	9.13 (0.24)	1.23 (0.27)	0.10 (0.07)	14.40 (7.27)

location overall performance when using covariance based estimates.

2) *Within $\hat{\Pi}$* : The first striking observation consists of the tremendous performance of the sparsity approach when applied to the estimation of the precision matrix: Volatility is remarkably low while Sharpe ratio is quite large. These performance are, however, achieved at the price of a higher turn-over rate and a smaller diversification level compared with the factor and shrinkage estimators. These observations hold irrespective of the estimation window size n .

Second, it is important to note that the shrinkage estimators

achieve very poor asset allocation performance when applied to the precision matrix. Factor based estimates display decent performance, but are however clearly and significantly outperformed by the sparsity-based estimators.

3) *Impact of short sale restrictions*: It is first striking to note the significant improvement of the performance of the sample covariance estimator $\hat{\Sigma}^{(1)}$ in the presence of short sale restrictions. It achieves levels of volatility and Sharpe ratios at par with the best estimators as soon as $n \gtrsim p$. Even in the singular case $n = 125$, the volatility of the optimal portfolio is much lower than in the absence of short-sale restrictions.

TABLE III
GLOBAL MINIMUM VARIANCE PORTFOLIO – 375 DAYS

375 days - Long/Short								
Σ				Π				
V	S	TO	H^{-1}	V	S	TO	H^{-1}	
Direct								
1	10.42 (0.26)	0.62 (0.26)	1.63 (0.58)	3.39 (1.66)	N/A	N/A	N/A	N/A
2	19.65 (0.51)	0.67 (0.26)	0.00 (0.00)	244.00 (0.00)	N/A	N/A	N/A	N/A
3	16.32 (0.37)	0.73 (0.25)	0.01 (0.01)	176.68 (16.84)	N/A	N/A	N/A	N/A
Factor								
4	10.52 (0.30)	1.20 (0.30)	0.08 (0.06)	27.50 (9.99)	10.75 (0.31)	1.18 (0.25)	0.08 (0.06)	26.29 (9.95)
5	10.38 (0.31)	1.15 (0.28)	0.07 (0.05)	31.74 (11.19)	10.60 (0.28)	1.08 (0.26)	0.05 (0.04)	56.61 (26.35)
6	9.49 (0.25)	1.35 (0.28)	0.11 (0.09)	25.38 (8.91)	N/A	N/A	N/A	N/A
Shrinkage								
7	8.70 (0.20)	0.96 (0.29)	0.86 (0.35)	7.19 (3.46)	19.22 (0.48)	0.68 (0.26)	0.04 (0.02)	230.12 (8.50)
8	8.30 (0.19)	0.96 (0.27)	0.82 (0.32)	6.48 (2.94)	18.06 (0.39)	0.60 (0.26)	0.04 (0.01)	145.92 (9.62)
Sparsity								
9	16.32 (0.38)	0.73 (0.30)	0.01 (0.01)	176.68 (16.84)	7.44 (0.18)	1.64 (0.29)	0.15 (0.09)	20.30 (7.32)
375 days - Long Only								
Direct								
1	9.20 (0.23)	1.21 (0.27)	0.09 (0.06)	10.83 (4.27)	N/A	N/A	N/A	N/A
2	19.65 (0.43)	0.67 (0.28)	0.00 (0.00)	244.00 (0.00)	N/A	N/A	N/A	N/A
3	16.35 (0.37)	0.73 (0.25)	0.01 (0.01)	177.47 (16.77)	N/A	N/A	N/A	N/A
Factor								
4	9.30 (0.24)	1.26 (0.26)	0.07 (0.05)	13.85 (5.48)	9.74 (0.23)	1.21 (0.26)	0.08 (0.06)	7.39 (3.30)
5	9.53 (0.24)	1.19 (0.27)	0.07 (0.05)	15.33 (6.14)	11.12 (0.30)	1.05 (0.24)	0.04 (0.03)	58.34 (31.11)
6	9.29 (0.22)	1.25 (0.28)	0.07 (0.06)	13.56 (4.93)	N/A	N/A	N/A	N/A
Shrinkage								
7	9.22 (0.20)	1.23 (0.26)	0.09 (0.06)	12.33 (4.71)	19.22 (0.46)	0.68 (0.27)	0.04 (0.01)	230.91 (7.31)
8	9.20 (0.24)	1.21 (0.27)	0.09 (0.06)	11.20 (4.44)	18.08 (0.47)	0.60 (0.28)	0.04 (0.01)	147.30 (8.93)
Sparsity								
9	16.35 (0.37)	0.73 (0.24)	0.01 (0.01)	177.47 (16.77)	9.24 (0.23)	1.31 (0.28)	0.07 (0.05)	15.15 (5.89)

TABLE IV
GLOBAL MINIMUM VARIANCE PORTFOLIO – 500 DAYS

500 days - Long/Short								
Σ				Π				
V	S	TO	H^{-1}	V	S	TO	H^{-1}	
Direct								
1	9.12 (0.23)	0.72 (0.27)	0.88 (0.35)	5.06 (2.30)	N/A	N/A	N/A	N/A
2	19.65 (0.44)	0.67 (0.25)	0.00 (0.00)	244.00 (0.00)	N/A	N/A	N/A	N/A
3	16.62 (0.47)	0.68 (0.28)	0.01 (0.01)	179.29 (15.69)	N/A	N/A	N/A	N/A
Factor								
4	10.91 (0.31)	1.13 (0.25)	0.06 (0.05)	27.40 (9.03)	11.10 (0.33)	1.12 (0.25)	0.06 (0.05)	26.49 (9.03)
5	10.72 (0.31)	1.10 (0.27)	0.06 (0.05)	31.71 (9.86)	11.08 (0.30)	0.96 (0.27)	0.04 (0.03)	65.18 (30.59)
6	9.85 (0.31)	1.34 (0.28)	0.08 (0.07)	25.69 (8.42)	N/A	N/A	N/A	N/A
Shrinkage								
7	8.55 (0.22)	0.89 (0.27)	0.65 (0.28)	7.43 (3.32)	19.06 (0.45)	0.65 (0.29)	0.04 (0.02)	219.21 (13.06)
8	8.25 (0.18)	0.94 (0.26)	0.63 (0.26)	6.98 (2.95)	18.11 (0.48)	0.57 (0.29)	0.03 (0.01)	146.69 (8.03)
Sparsity								
9	16.62 (0.45)	0.68 (0.26)	0.01 (0.01)	179.29 (15.69)	7.74 (0.18)	1.57 (0.31)	0.12 (0.07)	21.27 (7.23)
500 days - Long Only								
Direct								
1	9.37 (0.22)	1.13 (0.30)	0.07 (0.05)	11.11 (3.79)	N/A	N/A	N/A	N/A
2	19.65 (0.46)	0.67 (0.29)	0.00 (0.00)	244.00 (0.00)	N/A	N/A	N/A	N/A
3	16.64 (0.40)	0.68 (0.26)	0.01 (0.01)	180.05 (15.62)	N/A	N/A	N/A	N/A
Factor								
4	9.50 (0.21)	1.20 (0.25)	0.06 (0.04)	13.95 (4.64)	9.61 (0.22)	1.15 (0.23)	0.06 (0.05)	9.22 (3.48)
5	9.71 (0.25)	1.15 (0.26)	0.06 (0.04)	15.49 (5.13)	11.86 (0.31)	0.94 (0.28)	0.03 (0.02)	69.28 (33.62)
6	9.49 (0.27)	1.18 (0.28)	0.06 (0.05)	13.74 (4.25)	N/A	N/A	N/A	N/A
Shrinkage								
7	9.40 (0.20)	1.15 (0.25)	0.07 (0.05)	12.12 (4.05)	19.06 (0.42)	0.65 (0.27)	0.04 (0.01)	220.95 (11.50)
8	9.38 (0.23)	1.13 (0.31)	0.07 (0.05)	11.38 (3.89)	18.13 (0.47)	0.57 (0.25)	0.03 (0.01)	148.02 (7.42)
Sparsity								
9	16.64 (0.43)	0.68 (0.29)	0.01 (0.01)	180.05 (15.62)	9.50 (0.22)	1.26 (0.26)	0.05 (0.04)	15.88 (5.52)

In this respect, the results obtained in this study are perfectly in line with those reported in [27], which concluded that the introduction of constraints improves the optimization process when the sample covariance matrix is ill-conditioned.

Second, the differences in performance between the factor and shrinkage based estimators for the covariance matrix, which were observed in the absence of short sale restrictions are essentially smoothed out by the introduction of such constraints: Specifically, the performance in terms of volatility generally improve for the factor estimates while they slightly deteriorate for the shrinkage ones. The Sharpe ratios almost

never decrease; To the contrary, they increase for shrinkage estimators and reach the level achieved by the factor estimators. The turn-over and diversification also improve.

Finally, this regularization effect by short sale restrictions also operates for precision based asset allocation as, while still dominant, the sparsity based estimators outperform less significantly the factor based ones. Interestingly, when $n \leq p$, the sparsity based precision matrix estimator is the only one displaying satisfactory performance.

Elaborating on discussions in [27], the regularization effect induced by short sale restrictions can be understood as follows.

In the absence of short sale restrictions, Solution 3 to Problem 2 can be obtained via a classical Lagrangian formulation, $L(w, \lambda) = w' \Sigma w + \lambda (1_p' w - 1)$, (with λ a Lagrange multiplier), whose resolution is straightforward, yielding:

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

In the presence of further constraints induced by short sale restrictions, the Lagrangian formulation becomes $L(w, \lambda, \nu) = w' \Sigma w + \lambda (1_p' w - 1) - \nu' w$, (with λ and ν two Lagrange multipliers). Deriving the Karush-Kuhn-Tucker conditions of optimality yields

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

which, compared to the unconstrained solution, reads the same but for the covariance matrix Σ being replaced by a *shrunked effective* surrogate, $\Sigma + \text{Diag}(\nu) - \frac{1}{2} \cdot (1_p \nu' + \nu 1_p')$. 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.

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

While the shrinkage approach performs well when applied to the covariance matrix, it shows rather poor performance when applied to the precision matrix. This likely stems from the choice made for the target matrix in this case. Indeed considering, e.g., the shrinkage toward a diagonal matrix, the initial step consisted in retaining the diagonal of the sample precision matrix P_n as target, rather than the inverse of the diagonal of the sample covariance matrix: The inversion of S_n to obtain P_n already incorporates a significant amount of estimation noise at the initial step of the procedure.

The sparsity principle, applied to the covariance matrix performs rather poorly, while it yields the best performance when applied to the precision matrix. This can likely be explained by the structure of dataset studied here, which consists in shares from large capitalization European firms, that is in homogeneous assets, which, according to financial theory, are sensitive to a small number of common economic factors. They hence show little partial correlation and thus a sparse precision matrix.

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 factor covariance-based or sparsity precision-based asset allocation are on average globally equivalent, with however a slight yet clear advantage for this later class in volatility and Sharpe ratio, at the price of higher turnover and lower diversification. Notably for small n , $n \leq p$, sparsity precision-based asset allocations degrade in performance far less than other strategies and appear as the only viable option. 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.

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 extent, related to the so-called random matrix theory introduced in [49], has recently been brought back to the front of the scene in [30]. 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 [1], [18], [24], [47]. This is 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

$$L(\nu, \rho) = \text{E} \left[\|\rho \Delta + (1 - \rho) S_n - \Sigma\|^2 \right] \quad (27)$$

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 (27) with respect to Δ yields

$$\Delta = \text{Diag} \Sigma, \quad (28)$$

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

$$\hat{\Delta} = \text{Diag} S_n. \quad (29)$$

After substitution in (27), the minimization with respect to ρ leads to

$$\rho = 1 - \frac{\text{Tr}(\Sigma^2) - \text{Tr}[(\text{Diag} \Sigma)^2]}{\text{E}[\text{Tr}(S_n^2)] - \text{E}[\text{Tr}[(\text{Diag} S_n)^2]]}. \quad (30)$$

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 [40]

$$\text{Cov} \left((S_n)_{ij}, (S_n)_{kl} \right) = \frac{1}{n-1} \left(\begin{array}{c} \Sigma_{ik} \cdot \Sigma_{jl} \\ + \Sigma_{il} \cdot \Sigma_{jk} \end{array} \right). \quad (31)$$

As a consequence

$$\text{E}[\text{Tr}(S_n^2)] = \frac{n}{n-1} \text{Tr}(\Sigma^2) + \frac{1}{n-1} (\text{Tr} \Sigma)^2, \quad (32)$$

and

$$\mathbb{E} \left[\text{Tr} \left((\text{Diag} S_n)^2 \right) \right] = \frac{n+1}{n-1} \text{Tr} \left[(\text{Diag} \Sigma)^2 \right]. \quad (33)$$

Notice that these relations can straightforwardly be obtained by application of the Stein-Haff identity. By substitution of equations (32) and (33) in (30), we obtain the oracle shrinkage estimator

$$\rho = \frac{\text{Tr} (\Sigma^2) + (\text{Tr} \Sigma)^2 - 2 \text{Tr} \left[(\text{Diag} \Sigma)^2 \right]}{n \text{Tr} (\Sigma^2) + (\text{Tr} \Sigma)^2 - (n+1) \text{Tr} \left[(\text{Diag} \Sigma)^2 \right]}. \quad (34)$$

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

$$\hat{\Sigma}_j = \hat{\rho}_j \cdot \text{Diag} S_n + (1 - \hat{\rho}_j) \cdot S_n, \quad (35)$$

and

$$\hat{\rho}_{j+1} = \frac{\text{Tr} \left(\hat{\Sigma}_j \cdot S_n \right) + \left(\text{Tr} \hat{\Sigma}_j \right)^2 - 2 \text{Tr} \left[\left(\text{Diag} \hat{\Sigma}_j \right)^2 \right]}{n \text{Tr} \left(\hat{\Sigma}_j \cdot S_n \right) + \left(\text{Tr} \hat{\Sigma}_j \right)^2 - (n+1) \text{Tr} \left[\left(\text{Diag} \hat{\Sigma}_j \right)^2 \right]}. \quad (36)$$

By substitution of (35) into (36) we get

$$\hat{\rho}_{j+1} = \frac{1 - \phi_n \cdot \rho_j}{1 + (n-1) \cdot \phi_n - n \phi_n \cdot \rho_j}, \quad (37)$$

with

$$\phi_n = \frac{\text{Tr} (S_n^2) - \text{Tr} \left[(\text{Diag} S_n)^2 \right]}{\text{Tr} (S_n^2) + (\text{Tr} S_n)^2 - 2 \text{Tr} \left[(\text{Diag} S_n)^2 \right]}, \quad (38)$$

and $0 \leq \phi_n \leq 1$ by construction. Taking the limit as $j \rightarrow \infty$ 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

$$L(\nu, \rho) = \mathbb{E} \left[\|\rho \nu Id + (1 - \rho) P_n - \Pi\|^2 \right], \quad (39)$$

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

$$\hat{\nu} = \frac{1}{p} \text{Tr} P_n. \quad (40)$$

As a consequence, by substitution in (39) and minimization with respect to ρ we get

$$\rho = \frac{\mathbb{E} \left[\text{Tr} (P_n^2) \right] - \text{Tr} (\Pi^2) - \frac{1}{p} \text{Var} (\text{Tr} P_n)}{\mathbb{E} \left[\text{Tr} (P_n^2) \right] - \frac{1}{p} (\text{Tr} \Pi)^2 - \frac{1}{p} \text{Var} (\text{Tr} P_n)}. \quad (41)$$

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 $n > p + 2$)

$$P_n = \frac{n-p-2}{n-1} \cdot S_n^{-1} \quad (42)$$

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 [45], we know that

$$\text{Cov} \left((P_n)_{ij}, (P_n)_{kl} \right) = \frac{2 \cdot \Pi_{ij} \cdot \Pi_{kl} + (n-p-2) (\Pi_{ik} \cdot \Pi_{jl} + \Pi_{il} \cdot \Pi_{jk})}{(n-p-1)(n-p-4)}, \quad (43)$$

provided that $n > p + 4$. Hence

$$\mathbb{E} \left[\text{Tr} (P_n^2) \right] = \left(\frac{n-p}{(n-p-1)(n-p-4)} + 1 \right) \text{Tr} (\Pi^2) + \frac{n-p-2}{(n-p-1)(n-p-4)} (\text{Tr} \Pi)^2, \quad (44)$$

and

$$\text{Var} (\text{Tr} P_n) = 2 \cdot \frac{(\text{Tr} \Pi)^2 + (n-p-2) \cdot \text{Tr} (\Pi^2)}{(n-p-1)(n-p-4)}. \quad (45)$$

Thus, the oracle shrinkage parameter is

$$\rho = \frac{\frac{n-p-\frac{2}{p}(n-p-2)}{(n-p-1)(n-p-4)} \cdot \text{Tr} (\Pi^2) + \frac{n-p-2-\frac{2}{p}}{(n-p-1)(n-p-4)} \cdot (\text{Tr} \Pi)^2}{\left[1 + \frac{n-p-\frac{2}{p}(n-p-2)}{(n-p-1)(n-p-4)} \right] \text{Tr} (\Pi^2) + \left[\frac{n-p-2-\frac{2}{p}}{(n-p-1)(n-p-4)} - \frac{1}{p} \right] (\text{Tr} \Pi)^2}. \quad (46)$$

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

$$\hat{\Pi}_j = \hat{\rho}_j \cdot \frac{1}{p} \text{Tr} P_n \cdot Id + (1 - \hat{\rho}_j) \cdot P_n, \quad (47)$$

and

$$\hat{\rho}_{j+1} = \frac{\left[n-p-\frac{2}{p}(n-p-2) \right] \text{Tr} \left(\hat{\Pi}_j P_n \right) + \left(n-p-2-\frac{2}{p} \right) \left(\text{Tr} \hat{\Pi}_j \right)^2}{\left[n-p-\frac{2}{p}(n-p-2) + (n-p-1)(n-p-4) \right] \text{Tr} \left(\hat{\Pi}_j P_n \right) + \left[n-p-2-\frac{2}{p} - \frac{1}{p}(n-p-1)(n-p-4) \right] \left(\text{Tr} \hat{\Pi}_j \right)^2}. \quad (48)$$

The limit as $j \rightarrow \infty$ 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] Ben-Israel, A. and T.N.E. Greville (2003) *Generalized Inverses: Theory and Applications*. (Springer-Verlag New York, 2nd edition).
- [3] Bengtsson, C., and J. Holst (2003) On Portfolio Selection: Improved Covariance Matrix Estimation for Swedish Asset Returns. *Working paper* Lund University.
- [4] Bien, J., and R. J. Tibshirani (2011) Sparse Estimation of a Covariance Matrix. *Biometrika* **98**, 807–820.
- [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] Bouchaud, J.-P. and M. Potters (2009) *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management* (Cambridge University Press, 2nd Edition).

- [7] 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.
- [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] Dempster, A. P. (1972) Covariance Selection. *Biometrics* **28**, 157–75
- [17] Disatnik, D. J., and S. Benninga (2007) Shrinking the Covariance Matrix – Simpler is Better. *Journal of Portfolio Management* **33**(4), 55–63.
- [18] El Karoui, N. (2008) Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, **36**, 2757–2790.
- [19] 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.
- [20] 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).
- [21] Fama, E. F., and K. R. French (1992) The Cross-Section of Expected Stock Returns. *Journal of Finance* **47**, 427–465.
- [22] Friedman, J., H. Trevor, and R. Tibshirani (2008) Sparse Inverse Covariance Estimation with the Graphical Lasso. *Biostatistics* **9**, 432–441.
- [23] Haff, L. R. (1979) Estimation of the Inverse Covariance Matrix: Random Mixtures of the Inverse Wishart Matrix and the Identity. *Annals of Statistics* **7**, 1264–1276.
- [24] Haff, L. R. (1991) The Variational Form of Certain Bayes Estimators. *Annals of Statistics* **19**, 1163–1190.
- [25] Hero, A. O., and B. Rajaratnam (2011) Large Scale Correlation Screening. *Journal of the American Statistical Association* **106**, 1540–1552.
- [26] 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.
- [27] Jagannathan, R., and T. Ma (2003) Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *Journal of Finance* **58**, 1651–1684.
- [28] 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.
- [29] 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.
- [30] Laloux, L., P. Cizeau, J. P. Bouchaud, and M. Potters (1999). Noise Dressing of Financial Correlation Matrices. *Physical Review Letters* **83**, 1467–1470.
- [31] Ledoit, O. and S. Péché (2011) Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields* **151**, 233–264.
- [32] 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.
- [33] Ledoit, O., and M. Wolf (2012) Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices. *Annals of Statistics* **40**, 1024–1060.
- [34] Lian, H. (2011) Shrinkage Tuning Parameter Selection in Precision Matrices Estimation. *Journal of Statistical Planning and Inference* **141**, 2839–2848.
- [35] Malevergne, Y., and D. Sornette (2004) Collective Origin of the Coexistence of Apparent Random Matrix Theory Noise and of Factors in Large Sample Correlation matrices. *Physica A* **331**, 660–668.
- [36] Marcenko, V. A., and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics* **1**, 457–483
- [37] Markowitz, H. M. (1952). Portfolio Selection. *The Journal of Finance* **7**, 77–91.
- [38] Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments* (John Wiley & Sons).
- [39] Michaud, R. O. (1989) The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal? *Financial Analysts Journal* **45**, 31–42.
- [40] Muirhead, R. J. (2005) *Aspects of Multivariate Statistical Theory* (Wiley-Interscience, 2nd edition).
- [41] Ravikumar, P., M. Wainwright, G. Raskutti and B. Yu (2011) High-dimensional log-variance estimation by minimizing l_1 -penalized log-determinant divergence, *Electronic Journal of Statistics* **5**, 935–980.
- [42] Rothman, A. J. (2012) Positive definite estimators of large covariance matrices. *Biometrika* **99**, 733–740.
- [43] 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.
- [44] Sharpe, W. (1963) A Simplified Model for Portfolio Analysis. *Management Science* **79**, 277–231.
- [45] Siskind, V. (1972) Second Moments of Inverse Wishart-Matrix Elements. *Biometrika* **59**, 690–691.
- [46] 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.
- [47] Stein, C. (1975) Estimation of a Covariance Matrix. *Rietz Lecture, 39th Annual Meeting IMS*. Atlanta, Georgia.
- [48] Tanaka, M., and K. Nakata (2013) Positive definite matrix approximation with condition number constraint. *Optimization Letters*, DOI:10.1007/s11590-013-0632-7.
- [49] Wigner, E. P. (1953) On a Class of Analytic Functions from the Quantum Theory of Collisions. *Annals of Mathematics* **53**, 36–67.
- [50] Woerheide, W., and D. Persson (1993) An Index of Portfolio Diversification. *Financial Review Services* **2**, 73–85.
- [51] 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