ESTIMATING COVARIANCE MATRICES FOR PORTFOLIO OPTIMIZATION

GUILLAUME COQUERET AND VINCENT MILHAU

Abstract. We compare twelve estimators of the covariance matrix: the sample covariance matrix, the identity matrix, the constant-correlation estimator, three estimators derived from an explicit factor model, three obtained from an implicit factor model, and three shrunk estimators. Following the literature, we conduct the comparison by computing the volatility of estimated Minimum Variance portfolios. We do this in two frameworks: first, an ideal situation where the true covariance matrix would be known, and second, a real-world situation where it is unknown. In each of these two cases, we perform the tests with and without short-sales constraints, and we assess the impact of the universe and sample sizes on the results. Our findings are in line with those of Ledoit and Wolf (2003), in that we confirm that in the absence of short-sales constraints, shrunk estimators lead in general to the lowest volatilities. With long-only constraints, however, their performance is similar to that of principal component estimators. Moreover, the latter estimators tend to imply lower levels of turnover, which is an important practical consideration.

1. Introduction

Since the seminal work of Markowitz (1952) and Sharpe (1964), the Mean-Variance paradigm has been at the heart of Modern Portfolio Theory. Within this framework, the computation of efficient portfolios requires two key inputs: the vector of expected excess returns and the covariance matrix. Knowing these parameters, the investor is able to compute the set of efficient portfolios, that is the set of those portfolios that achieve the highest expected return for a given level of risk. Which efficient portfolio is selected depends on the amount of risk that the agent is willing to take, which is itself a function of his/her risk aversion. If short sales are allowed, the efficient frontier is generated by combining the minimum variance (MV) portfolio, which is the one chosen by investors with infinite risk aversion, and the maximum Sharpe ratio (MSR) portfolio. It is well known that the weights of these two strategies admit the following expressions:

$$\boldsymbol{w}^{MV} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{1_N}}{\boldsymbol{1_N'}\boldsymbol{\Sigma}^{-1}\boldsymbol{1_N}}, \tag{1}$$

and:

$$\boldsymbol{w}^{MSR} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{1_N'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}. \tag{2}$$

In these expressions, $\boldsymbol{\Sigma}$ is covariance matrix, $\boldsymbol{\mu}$ is the vector of expected excess returns, $\boldsymbol{1_N}$ is the N-dimensional vector of ones, and for each $k = 1, ..., N$, w_k^{MV} and w_k^{MSR} are the proportions of wealth which should be invested in asset k, in the MV and the MSR portfolio respectively.

A notorious problem with these strategies is that they rely on the unobservable parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$. Hence, they all involve *estimation risk*, in the sense that the estimated parameters differ from the true values. Plugging approximations into (1) or (2) therefore leads to non-optimal weights. In particular, the vector $\boldsymbol{\mu}$ is particularly hard to estimate. Numerous papers have shown that the use of historical averages for estimating expected returns implies poor out-of-sample performance (see Klein and Bawa (1976), Jobson and Korkie (1980), Merton (1980),

Date: December 20, 2013.
EDHEC Risk Institute.

Jorion (1985), Jorion (1991), Best and Grauer (1991), Chopra and Ziemba (1993), Britten-Jones (1999), Kondor et al. (2007)). The considerable difficulty to estimate $\boldsymbol{\mu}$ accurately has drawn interest for weighting schemes that do not use this parameter. These diversification schemes include maximum decorrelation, risk parity (Maillard et al., 2010) and maximum diversification (Choueifaty and Coignard, 2008). The corresponding portfolios can be computed with the sole knowledge of $\boldsymbol{\Sigma}$, which makes them of course insensitive to the estimation errors in $\boldsymbol{\mu}$. However, choosing one of these targets involves a loss in Sharpe ratio with respect to the choice of the MSR portfolio, unless $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ satisfy specific "optimality conditions".[1]

But estimation errors in the covariance matrix can be large too, if the universe is large and the sample available to estimate parameters is small: the ratio N/T, where N is the number of assets and T is the number of observations, plays an important role in determining the precision of the sample estimator (see Kan and Zhou (2007) and Fan et al. (2008) for rigorous proofs, as well as El Karoui (2010) and Kondor et al. (2007) for further theoretical and empirical insights on the subject). When the ratio N/T is large, Kan and Zhou (2007) even show that estimation errors in the sample covariance matrix contribute more than estimation errors in expected returns to the loss of efficiency of the proxied mean-variance efficient portfolio with respect to the true MSR portfolio. Furthermore, if N/T is less than one, the sample estimator is a singular matrix, which makes it impossible to compute a MV portfolio through (1). It is only when the sample size is large relative to the universe size (say, T is at least equal to $3N$), that the sample covariance matrix is suitable for portfolio optimization, as was shown in Pantaleo et al. (2011). In other situations, $\widehat{\boldsymbol{\Sigma}}$ should be regularized so as to mitigate the estimation risk. A review of regularization techniques proposed in the literature is presented in Section 2.

The purpose of this paper is to make a comparison between different estimators for the covariance matrix, which is the common input to MV, risk parity and maximum diversification portfolios. Specifically, we consider twelve estimators of the covariance matrix and their corresponding estimated MV portfolios:

- the sample estimator;
- the constant-correlation estimator, which assumes that all assets have the same pairs of correlation;
- the (scaled) identity matrix, which assumes that all assets are independent and have the same variance. The MV portfolio is then equally weighted;
- two single-factor estimators, in which the factor is taken to be either the return on an equally-weighted portfolio of all assets, or the return on a cap-weighted portfolio. These estimators are denoted with FEW and FCW;
- the three-factor estimator obtained from the Fama-French model, denoted with FFF;
- three shrunk estimators, that were introduced by Ledoit and Wolf in a series of paper in 2003 and 2004: they shrink the sample covariance matrix respectively towards the identity matrix, the equally-weighted single-factor matrix, and the constant-correlation matrix. These estimators are denoted with LWI, LWF and LWC;
- three estimators formed with principal component analysis, with different choices for the number of factors to be retained. These estimators are denoted with PCP, PCL and PCR.

These estimators are presented in details in Section 3. It should be noted that while inverting an estimator of $\boldsymbol{\Sigma}$ is the most obvious option to estimate minimum variance weights via (1), other techniques are available:

[1]The optimality conditions of a given portfolio are the conditions under which this portfolio coincides with the MSR portfolio. For instance, the MV portfolio equals the MSR one if, and only if, all assets have the same expected excess returns.

i) a direct estimation of $\mathbf{\Sigma}^{-1}$ (see Stevens (1998), Kourtis et al. (2012) and DeMiguel et al. (2013)), or *ii)* a direct estimation of the weights w_k^{MV} (see Clarke et al. (2013) and Frahm and Memmel (2010)). We leave these approaches outside the scope of this article, because they are too specific to the estimation of the MV portfolio, while an estimated covariance matrix can be used in many contexts: computation of other efficient portfolios or of "heuristic" portfolios such as risk parity and maximum diversification, computation of ex-ante variance or tracking error, etc.

Following the literature, we rank estimators by measuring the out-of-sample variance of estimated MV portfolios. We do this in two frameworks. In the first one, we simulate stock returns with a known covariance matrix, so that we are able to compute the ex-ante volatility of each portfolio, that is the volatility that would be achieved on average over an infinite number of paths. In the second situation, we use real financial data, which implies that the data-generating process and the true covariance matrix are unknown: we thus compute the out-of-sample volatility of each portfolio along the historical path. Because short-sales constraints are commonly imposed in practice, we consider long-only portfolios in addition to long-short ones. We also report the turnovers of the strategies: the turnover by itself is not an indicator of the accuracy of an estimator, but it is important for practitioners because it gives a sense of the magnitude of transaction costs.

Our main empirical findings can be summarized as follows:

(1) With simulated data, the ranking of estimators depends on the tails of the underlying distribution. If tails are thin, then two shrunk estimators (LWC and LWF) and one principal component estimator (PCP) display the best performances. But a fat-tail distribution is arguably a better representation of the observed distribution of stock returns. In this situation, one shrunk estimator (denoted with LWI) and another principal component estimator (PCR) stand out;
(2) Using real financial data, we obtain results consistent with those Ledoit and Wolf (2003): in a large universe (such as the S&P 500 universe) and assuming that short sales are allowed, the estimator shrunk towards the single-factor estimator implies an out-of-sample volatility which is lower than that associated with a principal component estimator (PCL), but the latter estimator leads to lower levels of turnover. We also recover results from Jagannathan and Ma (2003) regarding the impact of weight constraints;
(3) We complement this analysis by imposing short-sales constraints, which are often applied in practice. Under realistic implementation conditions (that is, computation of nonnegative weights by rescaling of the long-short weights), the constant-correlation estimator implies good results: it implies low out-of-sample volatilities and low levels of turnover. However, since this estimator has poor performances in the Monte-Carlo study with the Student distributions, we take these good results cautiously. But the principal component estimator denoted with PCR does not imply much higher volatility, while keeping the turnover reasonable.

Overall, the PCR estimator, in which the number of factors is chosen according to Random Matrix Theory (Marchenko and Pastur, 1967), displays good results in all tested situations. It is not the best in all of them, but the dominating estimators vary from one case to the other.

The remainder of the article is structured as follows. In Section 2, we review possible approaches to reducing estimation risk, in particular factor models and shrinkage techniques. In Section 3, we give a detailed description of the twelve estimators of the covariance matrix that we study in this paper, and which are representative of the available regularization methods surveyed in Section 2. Section 4 is devoted to the comparison of the estimators in the hypothetical situation where the true covariance matrix and the true asset return distribution are known. Section 5 provides a similar comparison, but uses real financial data, for which the true covariance matrix is unknown. In both cases, we will assess the estimation errors via the (ex-ante or out-of-sample) volatilities of estimated MV portfolios. We summarize our results and conclude in Section 6.

2. Review of regularization techniques for covariance matrix

The literature devoted to covariance matrix estimation is extensive. Recently, many new results and innovative approaches have blossomed, with applications to various fields (Finance, Statistics, Machine Learning, Genetics, etc.). A non exhaustive sample includes Bai and Shi (2011), Fan et al. (2011), Ledoit and Wolf (2012) and Mazumder and Hastie (2012), as well as the references therein. In this section, we wish to highlight the methods that have been developed with a view towards portfolio construction and admit an financial interpretation.

First, because $\widehat{\boldsymbol{\Sigma}}$ is a convergent estimator of $\boldsymbol{\Sigma}$, the sample estimator can be improved by increasing the sample size. This can be done either by considering older observations (going back further in time), or by using more frequent observations (weekly or daily returns instead of monthly or quarterly returns). However, long datasets are not always available for all stocks, and even if they were, their usage would raise questions as to the stationarity of parameters: indeed, the covariance structure may have varied over time, so that old data is not appropriate to estimate recent covariances. On the other hand, the use of daily data to estimate covariances has been shown to improve significantly the sample estimator (see e.g. Jagannathan and Ma (2003)). But high-frequency data raises specific issues, including asynchronous price problems as well as bid-ask bounces. In our studies below, we will therefore use weekly data.

Once the universe and the sample sizes have been specified, there are two alternative approaches which aim at reducing estimation risk. The first one focuses on the improvement of the input ($\widehat{\boldsymbol{\Sigma}}$), while the second one intends to regularize the outputs, that is, the estimated optimal weights. For instance, imposing constraints, such as bounds on the weights, can help to reduce the risk of optimal portfolios, as was shown in the seminal article of Jagannathan and Ma (2003). In this case, the MV portfolio is the solution of the following optimization program:

$$\left\{ \begin{array}{l} \min_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\ \text{s.t. } \boldsymbol{w}'\boldsymbol{1_N} = 1 \\ \quad \boldsymbol{lb} \leq \boldsymbol{w} \leq \boldsymbol{ub} \end{array} \right. , \tag{3}$$

where the second line represents the budget constraint (the portfolio must be fully invested) and the third line imposes the upper and lower bounds on the weights. Other robustification techniques include L^1 and L^2 norm constraints (DeMiguel et al., 2009) and shrinkage or mixing of weights (see Kan and Zhou (2007), Frahm and Memmel (2010), Tu and Zhou (2011) and Candelon et al. (2012))[2]. One advantage of these techniques is that they obviously help to address a notorious problem of $\boldsymbol{w}^{MV}$: the estimated minimum variance weights usually contain many large negative values. Even if such extreme short positions can be justified by the presence of a dominant factor in returns (Green and Hollifield, 1992), they are hardly implementable in practice. The regularizations contribute to reduce both the number and the magnitude of these negative weights.

[2]The L^1 norm of $\boldsymbol{w}$ is equal to the sum of absolute weights, $\sum_{i=1}^{N} |w_i|$, and its L^2 norm is equal to the square root of the sum of squared weights, $\sqrt{\boldsymbol{w}'\boldsymbol{w}}$.

In this article, we focus on the first approach, that is, the improvement of the estimated covariance matrix. When N/T is not close to zero, one of the main drawbacks of $\widehat{\mathbf{\Sigma}}$ is that it is widely dispersed: even though, on average, $\widehat{\mathbf{\Sigma}}$ is equal to $\mathbf{\Sigma}$ (the sample estimator is unbiased), there are often outliers for which $\widehat{\mathbf{\Sigma}}$ is very different from $\mathbf{\Sigma}$. These outliers are associated with large estimation risks and should be avoided. A reasonable compromise is to allow an estimator to have a nonzero bias in order to strongly reduce its variance. This trade-off between bias and variance is at the core of the two families of estimators that we present below: structured estimators and shrunk estimators.

The structured approach aims at reducing the dimensionality of the problem. The number of covariances to estimate within a universe of size N is $N(N+1)/2$, which is as a quadratic function of N. An option to reduce the number of parameters to estimate is to assume a constant correlation across stocks (the "overall mean model" of Elton and Gruber (1973)), so that the only parameters to be estimated are the volatilities. Factor estimators also alleviate the curse of dimensionality. Stock returns are assumed to be generated by a factor model, as in the Arbitrage Pricing Theory of Ross (1976):

$$R_t^{(i)} = \alpha^{(i)} + \boldsymbol{\beta}^{(i)\prime}\boldsymbol{F}_t + \varepsilon_t^{(i)}, \quad \text{for } i = 1, ..., N,$$

where $\boldsymbol{F}_t$ is a vector of K factor values, and the error terms are uncorrelated from the factors and uncorrelated across stocks. The covariance matrix is thus given by:

$$\mathbf{\Sigma} = \boldsymbol{\beta}'\mathbf{\Sigma}_F\boldsymbol{\beta} + \mathbf{\Sigma}_\varepsilon. \tag{4}$$

The number of parameters to estimate is thus: NK betas, plus $K(K+1)/2$ factor covariances, plus N idiosyncratic variances. It is linear in N, while the sample estimator of $\mathbf{\Sigma}$ contains $N(N+1)/2$ unrelated coefficients, which is a quadratic function of N. Imposing the factor structure on the estimator may introduce a bias, but will at the same time lower its variance because there will be fewer parameters to estimate.

The choice of the factors in this approach is of crucial importance. In the context of equity portfolios, usual models are the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and the three-factor Fama-French model introduced in Fama and French (1992). Factors are explicit in the sense that they are exogenously given, as opposed to being derived from the constituents' returns:[3] the market factor in the CAPM, and the size and the value factors in the Fama-French model. All of these models involve misspecification risk: in particular, the derivation of (4) strongly relies on the orthogonality between the residuals and the factors, a property that is not satisfied if relevant factors are omitted. On the contrary, implicit factor models consider factors which can be derived from the data only, without any assumption on the link between returns and exogenous factors. This avoids misspecification risk: for instance, with principal component analysis (PCA), the residuals from a K-factor model where K is the number of selected implicit factors are orthogonal to the factors by construction. A second advantage of implicit approaches is that they do not require other data than returns, while exogenous models do (for instance, fundamental data is needed to compute the Fama-French value factor). Another asset, which is specific to the PCA technique, is that the implicit factors are ranked according to their explanatory power with respect to historical returns. However, implicit approaches also have shortcomings: implicit factors may lack a financial interpretation, and the loadings of the principal component factors on the returns are often unstable.

The second family of robustification methods we are interested in relies on statistical shrinkage. It was first introduced in a financial setting by Barry (1974) for the Bayesian estimation of expected returns and was applied

[3]Exogenous factors are not necessarily observable. The classical example is the market portfolio of the CAPM (Roll, 1977).

to covariance matrix estimation by Frost and Savarino (1986) (see also Brown (1979)). The core idea of this technique is to combine two estimators so as to optimize the bias-variance trade-off. The first estimator should have zero bias but possibly a large variance while the second estimator must have a small variance, possibly at the cost of nonzero bias. The optimal linear combination of the two estimators will smaller bias than the second estimator and smaller variance than the first one. Consequently, the couple bias-variance will be more balanced. It is customary to take the sample covariance matrix as the first estimator. However, for the second estimator, no consensus exists among academicians or practitioners. A possible choice is the scaled identity matrix which may have a large bias but requires only one parameter to be estimated: the average variance of the stocks. Another popular choice is the single-factor estimator, defined as the estimator of the form (4) obtained by taking the average of stock returns as the single factor (Ledoit and Wolf, 2003). In this study, we also consider a cap-weighted version of the single factor, which is the usual proxy for the market portfolio of Sharpe's CAPM (Sharpe, 1964).

Table 1 summarizes the aforementioned regularization techniques as well as the related literature.

Improving the input (covariance matrix)		Regularizing the output (weights)
Factor Models	**Shrinkage**	
Explicit model (CAPM, Fama-French): Sharpe (1963) Fan et al. (2008)	Towards factor model covariance matrix: Ledoit and Wolf (2003)	Norm constraints: DeMiguel et al. (2009)
Implicit model (Principal component analysis): Laloux et al. (2000) Akemann et al. (2011) Chapter 40	Towards constant correlation covariance matrix: Ledoit and Wolf (2004a)	Lower and upper bounds: Jagannathan and Ma (2003)
	Towards identity matrix: Ledoit and Wolf (2004b) Sancetta (2008) Chen et al. (2011)	Shrinkage / Mixing: Kan and Zhou (2007) Frahm and Memmel (2010) Tu and Zhou (2011) Candelon et al. (2012)

TABLE 1. Summary of regularization methods studied in the literature.

3. SPECIFICATION OF THE COMPETING FAMILIES OF ESTIMATORS

3.1. **The benchmarks estimators.** In order to assess the performance of the enhanced estimators of $\mathbf{\Sigma}$ described below, we compare them to three simple benchmark estimators. The first one is the sample covariance. If $\boldsymbol{R}$ is the $(T \times N)$ matrix of returns and $\bar{\boldsymbol{R}}$ the $(1 \times N)$ vector of the stock's average returns, then the sample covariance matrix is defined as follows.

Definition 3.1. *The sample estimator, denoted by SAM is computed as*

$$\widehat{\mathbf{\Sigma}} = \frac{1}{T-1}\boldsymbol{R}'\boldsymbol{R} - \frac{T}{T-1}\bar{\boldsymbol{R}}'\bar{\boldsymbol{R}}.$$

This estimator can be used to compute a unique minimum variance portfolio as long as it is nonsingular: this condition is mathematically satisfied if, and only if, the universe size, N, is strictly smaller than the sample size, T. But in practice, having N close to T results in a ill-conditioned matrix, which is close to singular.

The second benchmark estimator is the constant correlation covariance matrix, which was introduced by Elton and Gruber (1973). In this case, the structure of the matrix is driven by the restriction that all pairwise correlations between assets are equal. In details, if we denote $\hat{\sigma}_{i,j}$ the elements of the sample covariance matrix and $\hat{c}_{i,j}$ those of the sample correlation matrix, the average sample correlation is computed as:

$$\hat{r} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{c}_{i,j}, \tag{5}$$

and the constant correlation covariance matrix is then given by:

$$\widehat{\boldsymbol{\Sigma}}^{CC} = [\hat{\sigma}^{CC}_{i,j}]_{ij}, \quad \text{with } \hat{\sigma}^{CC}_{i,j} = \begin{cases} \hat{\sigma}^{CC}_{i,i} = \hat{\sigma}_{i,i}, & i = j \\ \hat{\sigma}^{CC}_{i,j} = \hat{r}\sqrt{\hat{\sigma}_{i,i}\hat{\sigma}_{j,j}}, & i \neq j \end{cases} . \tag{6}$$

Thus, the estimator depends on the sample variances and the average correlation, that is $N+1$ parameters. Note that the realism of the constant correlation assumption depends on the universe: it is defensible within a single asset class, but not for portfolios encompassing many different asset classes.

Definition 3.2. *The constant correlation covariance matrix estimator, CC, will be computed as* $\widehat{\boldsymbol{\Sigma}}^{CC}$.

Lastly, for the sake of completeness and in order to verify that all estimators tend to bring an improvement over the situation where no parameter is estimated, we will test the identity matrix as an estimator for the covariance matrix. Using this estimator amounts to assuming that all correlations are zero and that all variances are equal to one. The assumption of unit variances may seem inappropriate, and one might instead propose the identity matrix multiplied by the average variance of stocks as an estimator. But for the purpose of computing MV portfolios, the value of the constant variance is unimportant, as can be seen from (1). Thus, we define the estimator as follows:

Definition 3.3. *The ID estimator of the covariance matrix is the scaled identity matrix.*

Of course, the use of this estimator leads to allocating equal weights to all constituents.

3.2. **Factor Models.**

3.2.1. *Introduction.* The expression for the covariance matrix in a multi-factor model has been given in Section 2. If we write $\hat{\boldsymbol{\beta}}$ for the $(N \times K)$ concatenation of all of the least-squares estimators of $\boldsymbol{\beta}^{(i)}$ (in the regression model (8)), then the factor estimator of the covariance matrix is

$$\widehat{\boldsymbol{\Sigma}}^{FAC} = \hat{\boldsymbol{\beta}} \widehat{\boldsymbol{\Sigma}}^{F} \hat{\boldsymbol{\beta}}' + \widehat{\boldsymbol{\Sigma}}_{\varepsilon}, \tag{7}$$

where $\widehat{\boldsymbol{\Sigma}}^{F}$ is the sample covariance matrix of the factors and $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$ is the diagonal matrix containing the variances of the residuals of the regression (8).

3.2.2. *Explicit factors.* We start by assuming that the factors have been determined exogenously, e.g. by an asset pricing model (CAPM or Fama-French). Choosing the CAPM, for instance, leads to the single-factor covariance estimator. In this setting, each asset's return is generated by the following linear model:

$$R^{(i)}_t = \alpha^{(i)} + \beta^{(i)} R^{(0)}_t + \varepsilon^{(i)}_t, \tag{8}$$

where the superscripts denote the asset indices (zero being the market factor) and the residuals $\varepsilon^{(i)}_t$ are uncorrelated to market returns and to one another. The covariance matrix implied by this model is

$$\widehat{\boldsymbol{\Sigma}}^{FAC} = \hat{v}^{(F)} \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}' + \widehat{\boldsymbol{\Sigma}}_{\varepsilon}, \tag{9}$$

where $\hat{v}^{(F)}$ is the sample variance of market returns, $\hat{\boldsymbol{\beta}}$ is the vector of the OLS estimators of the $\beta^{(i)}$ and $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ is the diagonal matrix containing the corresponding variances of the residuals. Note that $\widehat{\boldsymbol{\Sigma}}^{FAC}$ relies on the estimation of $2N+1$ numbers (N for $\hat{\boldsymbol{\beta}}$, N variances of residuals and $\hat{v}^{(F)}$), which is much smaller than $N(N+1)/2$ whenever the investment universe is large.

In practice, the market factor $R^{(0)}$ must be approximated by the returns on a stock index. We consider two proxies: the equally-weighted portfolio of the constituents, and the cap-weighted index, both of which are available on Kenneth French's website. This gives rise to two different estimators.

Definition 3.4. *The Equally-Weighted single-factor covariance matrix estimator, FEW, will be computed as* $\widehat{\boldsymbol{\Sigma}}^{FAC}$*, with the market factor being an equally-weighted portfolio of the constituents.*

Definition 3.5. *The Cap-Weighted single-factor covariance matrix estimator, FCW, will be computed as* $\widehat{\boldsymbol{\Sigma}}^{FAC}$*, with the market factor being a cap-weighted portfolios of all constituents.*

A refinement of the one-factor model is the three-factor Fama-French model. In Regression (8), the HML and SMB factors are added and the estimator takes the more general form (7), where $\boldsymbol{\beta}$ has dimension $(N \times 3)$.

Definition 3.6. *The Fama-French factor estimator for the covariance matrix, denoted as FFF, will be computed as in (7), where the betas are measured with respect to the three Fama-French factors.*

The use of this model is expected to bring an improvement over the use of a single-factor model, since three factors will mechanically explain a higher fraction of the total variance. On the other hand, it requires the estimation of more parameters.

3.2.3. *Implicit factors.* Implicit factors can be extracted from the data in numerous ways. We focus first on the extraction by eigendecomposition of the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$, that is the principal component analysis (PCA) of the matrix. By construction, the factors are orthogonal. They are usually sorted on their variances $\eta_1 \geq \cdots \geq \eta_N$, the η_i-s being the eigenvalues of the covariance matrix. In other words, the first factor relates to the largest eigenvalue and the last factor to the smallest one. Factor variances also have a statistical interpretation: the ratio of λ_i to the sum of eigenvalues, that is the ratio of λ_i to the sum of return variances, represents the fraction of total variance explained by the i^{th} factor. Hence the first factor is the one that explains the largest fraction of the total variance, while the latter one is merely noise. The selection procedure then reduces to choosing the appropriate number of factors: selecting a low number of factors reduces the explanatory power of the model, but retaining too many of them leads to retaining factors that are perhaps insignificant. This topic has generated a vast literature which is still expanding. Old tests (see Amemiya and Anderson (1990) for instance) tend to favor models with a high number of factors, but more recent criteria use regularization functions in order penalize models with large numbers of factors, as in Bai and Ng (2002).

The first principal component (PC) estimator we wish to test is based on this latter approach, which we detail now. We start by assuming that the PCA decomposition of the sample covariance matrix is known and that eigenvalues are ranked in decreasing order. Our objective is to obtain an optimal number of factors k^*. We perform N least-squares regressions of the form (8), where $\boldsymbol{F}_t$ is successively the vector of the first k factors, and k grows from 1 to N. We write $\boldsymbol{\varepsilon}_k$ for the corresponding vectors of $(T \times 1)$ residuals (we recall that T stands for the sample size). Bai and Ng (2002) propose the following criterion for k^*:

$$k^* = \underset{k}{\operatorname{argmin}} \left\{ \log(\boldsymbol{\varepsilon}_k' \boldsymbol{\varepsilon}_k) + k \left(\frac{N+T}{NT} \right) \log \left(\frac{NT}{N+T} \right) \right\}. \tag{10}$$

The idea is to minimize the sum of squared residuals with a penalty for high values of k. Originally, Bai and Ng (2002) propose to penalize simply the sum of squared residuals, $\varepsilon_k'\varepsilon_k$, but a scaling problem occurs, which is solved by taking the logarithm. For the sake of parsimony, we impose in our implementation of the criterion (10) that the integer k should not be greater than ten. Based on this criterion, we build our first principal component estimator.

Definition 3.7. *The PCP (for Principal Component selection based on Penalization) estimator of the covariance matrix is computed as* (7)*, where the number of factors is the* k^* *defined in* (10)*, subject to the restriction that* k *should be smaller than or equal to ten.*

The second PC estimator we investigate was proposed in Laloux et al. (2000); extensions can be found in Akemann et al. (2011), Chapter 40. The starting point is the eigendecomposition of the sample *correlation* matrix $\widehat{\mathbf{\Omega}}$:

$$\widehat{\mathbf{\Omega}} = \boldsymbol{P}\mathbf{\Lambda}\boldsymbol{P}', \tag{11}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ and $\boldsymbol{P}$ is an orthogonal matrix.[4]. Once again, the question is how many factors should be kept, in order to achieve a balance between explanatory power and statistical significance. An answer is provided by an elegant result in Random Matrix Theory (RMT), which characterizes the distribution of the eigenvalues when both N and T increase to infinity. This theorem, proved in Marchenko and Pastur (1967) states that when observations are independent and have the same stationary distribution with unit variance and N and the ratio N/T converges to a constant q different from 1, then, asymptotically, all nonzero eigenvalues of $\widehat{\mathbf{\Omega}}$ will lie in the interval

$$I = [(1-\sqrt{q})^2, (1+\sqrt{q})^2]. \tag{12}$$

The assumption that all assets have unit variance explains why this result is applied in the context of the eigendecomposition of the correlation matrix, not that of the covariance matrix. We also highlight the fact that it is an asymptotic result which holds only for N and T growing to infinity. In practice, it is applied for finite universes and samples, and the constant q is estimated as N/T. The upper bound of the interval I is estimated accordingly as $\nu_{N,T} = (1+\sqrt{N/T})^2$.

Surprisingly, when dealing with financial sample correlation matrices, a strong deviation from this theorem occurs: the largest eigenvalue lies considerably above the estimated theoretical upper bound, which is $\nu_{N,T}$. This can be explained by the fact that the investment universe is strongly driven by one factor, the "market factor", which accounts for a substantial proportion of the total variance of the universe (often more than 75% in the US equity market – see Brown (1989)). Depending on the number of assets under consideration, a few other eigenvalues may lie outside the theoretical bounds of the interval defined in (12). This discrepancy between the theoretical and the empirical distribution of eigenvalues is viewed as signal in Laloux et al. (2000). They consider that the gap between what is expected and what is observed is relevant information, while the conforming eigenvalues (i.e., those which lie within the theoretical bounds), are simply noise and should be discarded because they carry no useful information. Accordingly, they propose the following regularization of the covariance matrix.

Definition 3.8. *The PCR (for Principal Component selection based on RMT) estimator of the covariance matrix is computed by the following procedure:*

[4]This means that the inverse of $\boldsymbol{P}$ is the transpose of $\boldsymbol{P}$.

(1) *factorize the correlation matrix* $\widehat{\mathbf{\Omega}}$ *as* $\widehat{\mathbf{\Omega}} = \boldsymbol{P\Lambda P'}$*, with orthogonal* $\boldsymbol{P}$ *and diagonal* $\mathbf{\Lambda}$*,*

(2) *define the diagonal matrix* $\mathbf{\Lambda}^*$ *with the diagonal elements given by:* $\lambda_i^* = \lambda_i$ *if* $\lambda_i > \nu_{N,T}$*, and* $\lambda_i^* = 0$ *otherwise,*

(3) *compute the matrix product* $\widehat{\mathbf{\Omega}}^* = \boldsymbol{P\Lambda^* P'}$*,*

(4) *set the diagonal of* $\widehat{\mathbf{\Omega}}^*$ *to one:* $\hat{\rho}_{i,i}^* = 1$ *for* $i = 1, ..., N$*,*

(5) *finally, the PCR estimator of the covariance matrix is obtained by multiplying the robustified correlation matrix by the sample volatilities of assets:*

$$\widehat{\mathbf{\Sigma}}^* = \begin{bmatrix} \hat{\sigma}_{1,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \hat{\sigma}_{N,N} \end{bmatrix} \times \widehat{\mathbf{\Omega}}^* \times \begin{bmatrix} \hat{\sigma}_{1,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \hat{\sigma}_{N,N} \end{bmatrix}. \tag{13}$$

This robustification of $\widehat{\mathbf{\Sigma}}$ was first mentioned in Rosenow et al. (2002) and calls for a few remarks. First, Step *(2)* involves the choice of a degree of freedom, which is the common value assigned to the eigenvalues that lie below the bound $\nu_{N,T}$. Akemann et al. (2011) (Chapter 40) and Laloux et al. (2000) propose that the average value of the discarded eigenvalues, while we use zero in this article. In unreported results, we have checked that it makes only a marginal difference whether we choose zero or the average of the smallest eigenvalues (Pantaleo et al. (2011) also show empirically that the difference is particularly small when T is much larger than N). Second, Steps *(4)* and *(5)* will ensure not only that the final estimator preserves the sample variances, but also that it is invertible and that its eigenvalues are less dispersed than those of $\widehat{\mathbf{\Sigma}}$. Overall, this eigenvalue selection procedure will reduce the amplitude of the off-diagonal elements of $\widehat{\mathbf{\Sigma}}$, while maintaining its diagonal values. This effect is similar to that of bounds on weights (see Jagannathan and Ma (2003)): lower bounds shrink the largest covariances downwards, while upper bounds shrink the smallest values upwards. The PCR estimator is the one used by Amenc et al. (2011) as well as by Scientific Beta (http://www.scientificbeta.com) for the construction of smart beta indices that require the estimation of the covariance matrix.

Lastly, for the sake of completeness, we mention that other techniques have been developed in order to specify the number of factors to be retained. For instance, Connor and Korajczyk (1993) and Ahn and Horenstein (2013) propose alternative computations. However, the authors of the latter article point that their method does not apply in contexts where "a dominant factor (in terms of explanatory power) exists". The former article finds evidence for one to six factors in financial markets. In our study on out-of-sample performance, the PCR estimator selects one to seven factors, depending on the size of the investment universe, which varies from six to five hundred assets. Therefore, it seems that these two selection criteria give consistently close results.

Both the PCP and the PCR estimator allow the number of factors to fluctuate over time, as the PCA of the covariance or correlation matrix is performed on different windows. This is especially true for large investment universes. But one may be surprised by the assertion that the S&P 500 constituents would be driven by one or two factors during the 1960s and by five to seven factors during the 2000s. Thus, our last estimator is tailored so as to impose a time-invariant number of factors. Nevertheless, this number of factors should depend on the number of assets under consideration. If several studies have shown that it is likely that large equity universes are driven by approximately six factors (see Trzcinka (1986), Connor and Korajczyk (1993), Alessi et al. (2010) and in a slightly different context, Roll and Ross (1980)), it seems unlikely that six factors are required to model a ten-stock universe. We therefore propose the following heuristic threshold.

Definition 3.9. *The PCL (for Principal Component selection via the Logarithmic criterion) estimator is the factor estimator of the form* (7) *where the factors are the first k factors from the PCA of the covariance matrix, and k is the integer part of* $\log N$.

For instance, this criterion leads to retaining six factors for large universes, with five hundred of thousand assets, and two factors for a ten-stock universe.

Overall, PC estimators have the advantage of being computationally simple, because they require only a PCA factorization and a few simple matrix manipulations. However, it is not clear whether they are closer to $\mathbf{\Sigma}$ (for instance in the sense of quadratic distance) than the sample estimator. This is because the PC selection technique does not explicitly seek to minimize the distance between $\mathbf{\Sigma}$ and the proposed estimator.

3.3. **The shrinkage approach.**

3.3.1. *Introduction.* The second family of methods relies on a statistical shrinkage of $\widehat{\mathbf{\Sigma}}$. The rationale behind this technique is the following. The sample covariance matrix $\widehat{\mathbf{\Sigma}}$ has no bias but a lot of variance, therefore it seems appealing to combine it with an estimator with a nonzero bias, but little variance, known as the shrinkage target and denoted as $\vec{\mathbf{\Sigma}}$. In this case, the regularized estimator takes the form

$$\widehat{\mathbf{\Sigma}}^*_\alpha = (1-\alpha)\widehat{\mathbf{\Sigma}} + \alpha\vec{\mathbf{\Sigma}}. \tag{14}$$

The shrinkage target assumes a strong structure in the data and thus has a small variance, but it has possibly a large bias if the assumed structure is actually not present. The parameter α (shrinkage intensity) is chosen so as to ensure that the distance between $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}^*_\alpha$ is statistically as small as possible. Very often, the distance is assessed through a matrix norm $||\cdot||$[5] and the optimal intensity is defined by

$$\alpha^* = \underset{a}{\operatorname{argmin}}\ \mathbb{E}\left[\left\|\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}^*_a\right\|^2\right].$$

In practice, α^* depends on the true, unknown covariance matrix, and can only be estimated (the expressions for the estimators are given in the appendix). It also depends on the shrinkage target. In what follows, we present in details the three targets that we test in this paper, and which are borrowed from a series of papers by Ledoit and Wolf: the single-factor covariance matrix (Ledoit and Wolf, 2003), the scaled identity matrix (Ledoit and Wolf, 2004b) and the constant correlation covariance matrix (Ledoit and Wolf, 2004a).

3.3.2. *The three shrinkage estimators.* The first estimator takes the single-factor covariance matrix as the shrinkage target.

Definition 3.10. *The LWF estimator is the estimator of the form* (14) *in which the target is the factor estimator $\widehat{\mathbf{\Sigma}}^{FAC}$ defined in (9). The corresponding estimated intensity (α^*) is given in Ledoit and Wolf (2003) and is recalled in Appendix A.1.*

The second estimator takes the constant-correlation as shrinkage target.

Definition 3.11. *The LWC estimator is the estimator of the form* (14) *in which the target is the constant-correlation matrix $\widehat{\mathbf{\Sigma}}^{CC}$. The corresponding estimated intensity (α^*) is given in Ledoit and Wolf (2004a) and is recalled in Appendix A.2.*

[5] For instance, the spectral, Frobenius or max norm.

The last shrinkage target is simply a multiple of the identity matrix. In order for $\widehat{\boldsymbol{\Sigma}}$ and $\vec{\boldsymbol{\Sigma}}$ to have a comparable scale, we set in this case $\vec{\boldsymbol{\Sigma}} = \bar{\sigma}^2 \boldsymbol{I}_N$, where $\boldsymbol{I}_N$ is the identity matrix of size N, and $\bar{\sigma}^2$ is the average of sample variances.

Definition 3.12. *The LWI estimator is the estimator of the form* (14) *in which the target is the scaled identity matrix. The corresponding estimated intensity is given in Ledoit and Wolf (2004b) and is recalled in Appendix A.3.*

3.3.3. *Comments.* The estimators LWF, LWI and LWC have become references in the literature on portfolio optimization. For instance, in the past few years, the shrinkage towards the single factor covariance matrix was used as benchmark in the following articles: Barras (2007), DeMiguel et al. (2009), Liu (2009), Behr et al. (2012), Santos et al. (2012) – a list which is far from exhaustive.

We will therefore consider the above three shrinkage schemes for $\widehat{\boldsymbol{\Sigma}}$. Since the seminal article of Ledoit and Wolf (2003), several extensions have blossomed (see e.g. Sancetta (2008), Chen et al. (2011), Candelon et al. (2012), to name but a few). To the best of our knowledge, no extensive comparative study has yet demonstrated the superiority of these extensions over the original methods of Ledoit and Wolf in the field of portfolio optimization. We will therefore focus on these three estimators.

Shrinkage approaches have connections with other regularization techniques. First, the shrinkage towards the scaled identity matrix (LWI) amounts to shrinking the eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ towards their grand mean and hence reduces their dispersion. Therefore, the effect of this particular method is similar to that of Step *(4)* in Definition 3.8. Second, Jagannathan and Ma (2003) proved that the solution to the variance minimization problem (3) is equal to the solution of the same problem where the bounds on weights are removed and the covariance matrix $\boldsymbol{\Sigma}$ is replaced by a shrunk version of $\boldsymbol{\Sigma}$ (specifically, low covariances are increased, and high covariances are decreased). Third, if norm constraints are used instead of bound constraints, then DeMiguel et al. (2009) show that the optimal weights can also be interpreted as the solution to an unconstrained variance minimization problem where the sample covariance matrix is replaced by a shrunk estimator. These examples show that shrinkage estimators encompass many forms of regularizations.

The shrinkage estimators have two main advantages. First, for standard shrinkage targets, such as those listed above, the optimal intensity can be approximated with a closed-form expression, and no optimization is required to estimate it. Second, these estimators are explicitly designed to minimize the distance from the true matrix within the family of estimators of the form (14). In this sense, they can be considered as optimal, which was not the case for the PC estimators. But a drawback of shrinkage estimators is that they depend on the choice of a shrinkage target, which is after all arbitrary.

We are now equipped with twelve estimators: three benchmarks (SAM, CC and ID), three shrinkage estimators (LWF, LWC and LWI), three estimators based on the PCA of the sample covariance or correlation matrix (PCL, PCP and PCR), and three estimators based on explicit factor models (FEW, FCW and FFF). Note that since the first factor is usually close to an equally-weighted portfolio of the assets, the FEW estimator is comparable to a PC estimator with only one factor. The remainder of the article is devoted to the comparison of these estimators in various empirical settings.

4. Comparison with simulated data (known $\boldsymbol{\Sigma}$)

In our first setup, we wish to quantify the estimation risk for each of the ten estimators (the FCW and FFF estimators cannot be computed in this setting), while assuming that $\boldsymbol{\Sigma}$ is known. For this purpose, we use Monte-Carlo simulations and generate paths for asset returns with a given covariance matrix. This will enable us to know the true MV weights and the corresponding ex-ante volatility, that is, the lowest possible volatility which can be attained by a portfolio. A good performance indicator of an estimator is then the increase between this true minimal volatility and that of the estimated MV portfolio. In this context, one of the main conveniences of the Monte-Carlo setup is that it allows to control the distribution of returns. We wish to take advantage of this flexibility to test the impact of distribution tails on our results.

We do not include in this study the estimators FCW and FFF because they rely on the CAPM or Fama-French model and hence require paths for either the market factor or the three Fama-French factors. Thus, only the three benchmark estimators (SAM, CC and ID), the three principal component estimators (PCL, PCP and PCR) and the three shrunk estimators (LWC, LWF and LWI) are considered here. The FEW is also provided and the single factor in this case is simply an equally-weighted portfolio of all of the assets.

4.1. **The Monte-Carlo protocol.** We compare the nine estimators on a population of true covariance matrices. In order to generate realistic matrices, we use the Fama-French model. Specifically, we divide the period from 1957 to 2010 into 527 two-year windows of weekly returns, the starting dates of two consecutive windows being separated by five weeks. In each window, we consider the N stocks that have stayed within the S&P 500 universe, and we estimate their Fama-French betas by running the following linear regression:

$$R_t^{(i)} - R_t^{(f)} = \alpha^{(i)} + \sum_{k=1}^{3} \beta_k^{(i)} F_t^{(k)} + \epsilon_t^{(i)}, \quad i = 1, \dots, N, \tag{15}$$

where $R_t^{(f)}$ is the value of the risk-free rate at the beginning of period t. We then construct the true covariance matrix as:

$$\boldsymbol{\Sigma} = \boldsymbol{\beta}' \boldsymbol{\Sigma}_F \boldsymbol{\beta} + \boldsymbol{\Sigma}_\epsilon,$$

where $\boldsymbol{\Sigma}_F$ is the (3×3) sample covariance matrix of the Fama-French factors, $\boldsymbol{\beta}$ is the $(3 \times N)$ matrix of factor loadings estimated over the two-year window and $\boldsymbol{\Sigma}_\epsilon$ is the $(N \times N)$ diagonal matrix containing the variances of the residuals. We also compute the true expected excess returns as:[6]

$$\mu^{(i)} = \alpha^{(i)} + \sum_{k=1}^{3} \beta_k^{(i)} m^{(i)},$$

where $\alpha^{(i)}$ is the estimate for the intercept in Regression (15) and $m^{(k)}$ is the risk premium associated with the k^{th}-factor. The risk premia are estimated as the long-term averages of the returns of the Fama-French factors.

Once the covariance matrix and expected returns are computed for the whole S&P 500 universe, we form two sub-universes, in order to study the effect of universe size on our results and to have a constant size over time (indeed, the number of constituents of the S&P 500 index slightly varies over time). These two universes consist respectively of fifty and hundred stocks. The selection of stocks in performed in the following way. For each window, the stocks are sorted on their true expected return $\mu^{(k)}$, and the fifty and hundred stocks are chosen uniformly according to the distribution of the population's expected returns (that is, one pick every five or ten stocks after sorting). Once this selection is completed for a given window i, we compute the true covariance matrix and the true vector of expected returns, $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\mu}_i$, for the selected subset of stocks (in the sequel, i will stand as the index of the 527 known covariance matrices). At this stage, we have a population of 527 pairs of "true" values $(\boldsymbol{\Sigma}_i, \boldsymbol{\mu}_i)$.

---

[6]In what follows, we refer to expected excess returns simply as expected returns.

The next step is to generate estimates for $\mathbf{\Sigma}_i$. To this end, we assume that the covariance matrix is estimated as its sample counterpart from a sample of T observations. To study the effect of sample size, we consider different values of T, namely $T = 26$, 52, 104 or 208 weeks. For each sample size, we simulate random paths for asset returns with covariance matrix $\mathbf{\Sigma}_i$ and vector of expected values $\boldsymbol{\mu}_i$. The impact of the distribution tails will be assessed by comparing the results obtained for a multivariate Gaussian distribution with those obtained for a multivariate Student distribution with three degrees of freedom (this choice is arbitrary but ensures that the distribution has both heavy tails and finite variance). We underline that in both cases, the distributions will have common mean $\boldsymbol{\mu}_i$ and covariance matrix $\mathbf{\Sigma}_i$.[7] Based on each simulated sample j, where j varies from 1 to 50, we compute the sample covariance matrix $\widehat{\mathbf{\Sigma}}_{i,j}$, as well as the other eight estimators of $\mathbf{\Sigma}_i$. The corresponding MV weights (1) are obtained as:

$$\widehat{\boldsymbol{w}}_{i,j}^{MV} = \frac{\widehat{\mathbf{\Sigma}}_{i,j}^{-1}\mathbf{1}}{\mathbf{1}'\widehat{\mathbf{\Sigma}}_{i,j}^{-1}\mathbf{1}}. \tag{16}$$

To compare the nine estimators, we compute the "Relative Increases in ex-ante Volatility", defined as:

$$RIV(\boldsymbol{w}_i^{MV}, \widehat{\boldsymbol{w}}_{i,j}^{MV}, \mathbf{\Sigma}_i) = \frac{\hat{\sigma}_{i,j}^{MV} - \sigma_i^{MV}}{\sigma_i^{MV}}, \quad i = 1, ..., 527, \quad j = 1, ..., 50. \tag{17}$$

The ex-ante volatilities of the true and estimated MV portfolios are computed as:

$$\sigma_i^{MV} = \sqrt{(\boldsymbol{w}_i^{MV})'\mathbf{\Sigma}_i \boldsymbol{w}_i^{MV}}, \quad \hat{\sigma}_{i,j}^{MV} = \sqrt{\left(\widehat{\boldsymbol{w}}_{i,j}^{MV}\right)' \mathbf{\Sigma}_i \left(\widehat{\boldsymbol{w}}_{i,j}^{MV}\right)}.$$

The ex-ante volatility of the true MV portfolio is the minimum possible volatility, that would be obtained if the weights were computed with perfect knowledge of $\mathbf{\Sigma}_i$. In practice, of course, the true covariance matrix is not known and is estimated, so that the estimated MV portfolio has a higher ex-ante volatility, $\hat{\sigma}_{i,j}^{MV}$. The RIV measures the increase in volatility incurred by the use of the estimator as opposed to the true covariance matrix.

In general, the portfolio (16) has some negative weights. But short positions are not realistic, because many funds or investment vehicles have long-only policies and long-short portfolios have in general higher turnover, and thus higher transaction costs, than long-only versions: long-short MV weights often display instability, and they require substantial adjustments at rebalancing dates, which can be highly costly. Thus, we also compute an estimated long-only MV portfolio, by solving the following program:

$$\min_{\substack{\boldsymbol{w} \\ \boldsymbol{w} \geq 0}} \sqrt{\boldsymbol{w}'\mathbf{\Sigma}_{i,j}\boldsymbol{w}}. \tag{18}$$

The corresponding portfolio is denoted as $\widehat{\boldsymbol{w}}_{i,j}^{MV-LO}$. One drawback of long-only solutions is that they admit no closed-form expression as functions of the covariance matrix. Thus, they must be numerically computed, which is time-consuming, especially for the larger universe, that contains hundred assets. An alternative approach to computing long-only portfolios is the following "rescaling" procedure, that adjusts the unconstrained weights in

[7]Note that this requires the multiplication of $\mathbf{\Sigma}_i$ by the scaling factor $(\nu - 2)/\nu$ in the simulations of the Student distribution to ensure that returns do have $\mathbf{\Sigma}_i$ as covariance matrix. The number of degrees of freedom in this case is $\nu = 3$.

order to obtained nonnegative weights:

$$\text{(19)} \qquad \begin{cases} \text{1 - compute the optimal unconstrained weights } \widehat{\boldsymbol{w}}_{i,j}^{MV}; \\ \text{2 - set the negative weights to zero;} \\ \text{3 - normalize the positive weights so that they sum to one ;} \\ \text{4 - denote the resulting weights as } \widehat{\boldsymbol{w}}_{i,j}^{MV-R}. \end{cases},$$

This procedure has the advantage of being computationally simple, in that it requires no numerical optimization. However, it is not optimal, since it does not in general lead to the same weights as the variance minimization subject to nonnegativity constraints. In the end, we gather, for each regularization scheme, 527×50=26350 performance measures. We provide their averages in Table 4 (Panel A). Moreover, for a fixed $\boldsymbol{\Sigma}_i$ we compute the standard deviation of the RIV across the fifty noisy estimates, and then average these standard deviations over all true covariance matrices. These average standard deviations are collected in Panel A of Table 5.

As for long-short portfolios, we compute the relative increases in ex-ante volatilities for the two long-only portfolios as:

$$\text{(20)} \qquad RIV(\boldsymbol{w}_i^{MV-LO}, \widehat{\boldsymbol{w}}_{i,j}^{MV-LO}, \boldsymbol{\Sigma}_i) = \frac{\sqrt{(\widehat{\boldsymbol{w}}_{i,j}^{MV-LO})'\boldsymbol{\Sigma}_i \widehat{\boldsymbol{w}}_{i,j}^{MV-LO}} - \sqrt{(\boldsymbol{w}_i^{MV-LO})'\boldsymbol{\Sigma}_i \boldsymbol{w}_i^{MV-LO}}}{\sqrt{(\boldsymbol{w}_i^{MV-LO})'\boldsymbol{\Sigma}_i \boldsymbol{w}_i^{MV-LO}}},$$

$$\text{(21)} \qquad RIV(\boldsymbol{w}_i^{MV-LO}, \widehat{\boldsymbol{w}}_{i,j}^{MV-R}, \boldsymbol{\Sigma}_i) = \frac{\sqrt{(\widehat{\boldsymbol{w}}_{i,j}^{MV-R})'\boldsymbol{\Sigma}_i \widehat{\boldsymbol{w}}_{i,j}^{MV-R}} - \sqrt{(\boldsymbol{w}_i^{MV-LO})'\boldsymbol{\Sigma}_i \boldsymbol{w}_i^{MV-LO}}}{\sqrt{(\boldsymbol{w}_i^{MV-LO})'\boldsymbol{\Sigma}_i \boldsymbol{w}_i^{MV-LO}}}.$$

Note that in (20), the increase in volatility is only caused by the estimation error, while in (21), the increase in volatility is also due to the non-optimality of the solution.

4.2. **Discussion of the results.** Average values and average standard deviations of RIV for the various estimators are shown in Tables 4 and 5.

4.2.1. *Impact of universe and sample sizes and distribution tails.* We first observe that the (average) RIV is a decreasing function of the sample size: for a fixed N, the ex-ante volatility decreases as T increases. This happens because a larger sample allows for a more accurate estimation of the true covariance matrix. The reduction is particularly strong for long-short MV portfolios constructed from the sample estimator: with small samples, the sample covariance matrix is not even invertible, so that the corresponding MV portfolio is not uniquely defined, and no value is reported in the table. The ex-ante volatility of the estimated MV portfolio is then rapidly decreasing in the sample size. The same qualitative effect is observed for all other estimators (the CC one, the PC ones, the LW ones and the FEW one), but the magnitude is not as large. For the ID estimator, of course, the sample size is irrelevant since the sample is actually not used. We also note that the RIV is increasing in the number of assets: for a given sample size and estimation method, the average RIV is always larger with hundred than with fifty constituents. In view of these results, it is not surprising that the RIV is minimal when the ratio N/T is as small as possible. In the table, the minimum ratio corresponds to a fifty-stock universe with a sample of 208 weeks: it can be checked that for each estimation method and each distribution of returns, this combination leads to the lowest RIV. But the RIV is not only a function of the ratio N/T. For instance, in each panel, a comparison between the lines corresponding to $N = 50$ and $T = 104$ and those corresponding to $N = 100$ and $T = 208$ (e.g., lines 3 and 8 and lines 11 and 16 in Panel A) shows that for most estimation methods, the RIV is higher if $N = 100$, although the ratio N/T is unchanged.

We next look at the impact of tails. The Gaussian distribution has zero kurtosis, while the Student distribution with three degrees of freedom has an infinite kurtosis. It turns out that for given universe size, sample size and estimation method, the RIV is much larger when the distribution has fat tails than when it is Gaussian: the ratio of the value achieved with the Student distribution to the value achieved with Gaussian returns is in general comprised between two and three. Hence, estimation errors are magnified if asset returns have fat tails. This is an important property because the historical distribution is known to exhibit fat tails (we compare the estimators on historical data in Section 5). Thus, the volatility of the estimated MV portfolio tends to be larger if the universe is large, the sample is small and the underlying return distribution has non-zero kurtosis.

4.2.2. *Best and worst estimators for given universe and sample sizes and given distribution of returns.* The next natural question is which estimators perform best or worse when these three dimensions have been fixed. A first observation is that the SAM estimator (i.e., the sample covariance matrix) leads to the highest RIV for almost every combination of the three criteria. There are however six exceptions, on lines 28, 31, 32, 43, 44 and 48, where this estimator dominates the CC and the LWC ones. These exceptions correspond to the largest sample sizes (104 or 208 weeks) and to long-only portfolios (computed either in an optimal way, as in Panel B, or in a heuristic way, as in Panel C). This is in line with the findings of Jagannathan and Ma (2003), who show that increasing the number of observations or imposing weight constraints reduces the out-of-sample variance of a MV portfolio based on the SAM estimator. More generally, the difference between the SAM estimator and the others is extremely large when short positions are allowed and the sample is small (line 2), but reduces if one goes for a larger sample or if one imposes nonnegativity constraints, even in a non-optimal way (lines 12, 20, 36 for instance). The CC estimator does not systematically dominate the SAM one (lines 31 and 43 for instance). This is not surprising for large samples, because the assumption of a uniform correlation structure is less realistic with hundred than with fifty constituents, but the CC estimator can be worse even in small samples (line 28).

Turning to factor-based and shrunk estimators (i.e., the PC, LW and FEW ones), we observe that in all cases, at least one of them outperforms the three benchmark estimators. In other words, the estimator that yields the lowest volatility is always to be searched for among those that rely on a factor model or on statistical shrinkage, which confirms the interest of these regularization methods. That is not to say, however, that a given structured estimator always outperforms the benchmarks. In details, we see that the PC estimators outperform the benchmark ones more often than the LW ones do. Indeed, there are only nine lines where at least one PC estimator underperforms one of the benchmark ones,[8] while there are thirty-four lines in which at least one LW estimator is dominated by a benchmark estimator are in much larger number.[9] The two LW estimators that most often underperform a benchmark estimator are LWC and LWI. The FFF estimator (i.e., the one that relies on the Fama-French model) outperforms the three benchmarks in the majority of cases, but it does it less often than the PC estimators. Another observation is that the choice of the shrinkage target has more impact on the RIV of the LW estimators than the factor selection criterion used in the PC estimators. This can be seen with the spread between the highest and the lowest RIV within a class of estimators: it is only 2.3% for PC estimators (line 25), and it reaches 24% for the LW ones (line 13).

[8]These lines are 1, 2 (PCL is dominated by CC), 9, 19, 11 (PCL, PCP and PCR are dominated by ID), 17 (PCL is dominated by CC), 18 (PCL, PCP and PCR are dominated by CC), 25 (PCL is dominated by ID) and 26 (PCP and PCR are dominated by ID)

[9]The complete list is 1, 2, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 26, 27, 28, 30, 31, 32, 33, 34, 35, 37, 38, 39, 41, 42, 43, 44, 45 and 48, which represents 34 lines out of 48.

Looking at the minimum RIV in each row, we see that the lowest RIV is attained with the PCP, the LWC or the LWF estimator if the underlying distribution is Gaussian and portfolios are computed by numerical optimization, either without or with short-sales constraints. If weights are rescaled, the minimum can be achieved by the FEW estimator, but the lowest RIV achieved with the PCP, LWC and LWF estimators exceeds this minimum by no more than 30 basis points (lines 37, 38 and 39). But for practical applications, the results obtained with the Student distribution may be more relevant, given the presence of fat tails in historical return distribution. In these cases, the minimum is almost always attained by the LWI estimator, except in lines 16, 32, 47 and 48, where it is the PCR estimator that dominates the others. When the minimum is achieved by the LWI estimator, the RIV of the PCR estimator exceeds this minimum by no more than 650 basis points. This number may seem huge, but it should be noted that by definition of the RIV (see (17), (20) and (21)), an increase of 650 basis points in the RIV corresponds to an increase in the volatility of the estimated MV portfolio which is equal to 0.0650 times the volatility of the true MV portfolio. Assuming that the latter volatility is between 15% and 20%, this increase in RIV means an increase in volatility comprised between 0.98% and 1.30%, which is reasonable.

Looking at the average standard deviations of Table 5, we see that in the presence of heavy tails and when the universe is larger than the available sample, the LWC, LWF and CC may produce outliers, that is, weighting schemes for which the ex-ante volatility is substantially (often more than 40%) larger than that of the true MV portfolio.[10]

4.2.3. *Impact of rescaling.* As a last comment on Table 4, we observe that for fixed universe and sample sizes, estimation method and return distribution, the average RIV in Panel C can be lower than that in Panel B. This result is perhaps surprising, given that the rescaling is a non-optimal way of computing long-only MV weights. But because ex-ante volatilities are computed using the true covariance matrix, not the estimated one, there is no mathematical reason for the portfolio with rescaled weights to have a larger variance. The fact that this portfolio has sometimes a lower variance means that the estimation errors compensate the loss of optimality which arises from the use of a heuristic rule.

5. Comparison with real financial data (unknown Σ)

5.1. **The empirical protocol.** We now wish to assess the quality of the twelve estimators on live financial data. In this case, the true covariance matrices are unknown and it is impossible to compute the RIV as in the previous setting. A standard empirical procedure is to estimate MV portfolios and measure their out-of-sample volatility.

In details, we consider the seven investment universes presented in Table 3. Repeating the analysis on several datasets, as in DeMiguel et al. (2009), ensures that results are not driven by the specific data pattern encountered in a universe. We also test different sample sizes: the covariance matrix is estimated using T weekly returns, with T being equal to 52, 104 or 208 weeks. Then, for each estimator, we compute three estimated MV portfolios, as in the previous section: unconstrained (long-short), optimal long-only, and long-only with rescaled weights. The portfolios are rebalanced every thirteen weeks, that is, four times a year. Our primary focus is set on the out-of-sample volatility, that is, the standard deviation of the returns over the whole period.[11] We have

[10] Lines 9, 13, 25 and 29 show that the average standard deviation in *RIV* of LWC, LWF or CC estimators can be twice as large as those of LWI or PCR estimators.

[11] The first T weeks in the dataset are not used in the backtest because they are needed for the estimation of the first covariance matrix.

also computed the turnover of each of the four strategies, as it is a practical indicator which gives a sense of the magnitude of transaction costs. For both indicators, we use the same definitions as DeMiguel et al. (2009) (Equations (11) and (13) in their paper). In particular, the turnover on rebalancing date t is computed as:

$$\text{Turnover}_t = \sum_{i=1}^{N} |w_{it} - w_{it-}|,$$

where w_{it} is the imposed weight of constituent i on date t and w_{it-} is the effective weight just before date t. The turnovers that we report in the tables below are thus quarterly.

5.2. **Discussion of the results.** The volatilities and the one-way turnovers of long-short portfolios are in Tables 6 and 7. Those for the MV portfolio subject to long-only constraints (solutions to (18)) are in Tables 8 and 9. Finally, Tables 10 and 11 display the statistics for the MV portfolios with rescaled weights.

5.2.1. *Comparison with the literature.* A first remark is that our results are consistent with those of Ledoit and Wolf (2003), in spite of several differences between their protocol and ours: their study is carried on a very large investment universe (more than nine hundred stocks), while our largest universe (the S&P 500 universe) has five hundred stocks, they estimate the covariance matrix from monthly returns while we use weekly observations, and they rebalance portfolios every year as opposed to every quarter. Besides, their principal component estimator retains five factors from the decomposition of the covariance matrix, while we diagonalize the correlation matrix and test different criteria for factor selection (see Section 3.2). The number of of five factors is close to the number of six which is given by the Logarithmic criterion for the S&P 500 universe ($\log(500) = 6.21...$). Finally, Ledoit and Wolf (2003) do not impose short-sales constraints. Thus, the results that they report in their Tables 1 and 2 can be compared with those that we obtain in our Table 6 for the S&P 500 universe, and in particular, their principal component estimator is roughly the same as our PCL estimator. Ledoit and Wolf (2003) show that the LWF estimator leads to the lowest volatility, which is also the case for all sample sizes in Table 6. They also find that the LWI estimator is the second best in terms of out-of-sample volatility. In the column SPX of Table 6, we actually find that the second best estimator is PCL or PCR (the LWI estimator turns out to be the second best when we average volatilities across datasets). We also confirm that the PCL estimator yields slightly higher volatilities than the LWF one, but the increase with respect to LWF is not huge (the largest spread is 60 basis points, with $T = 208$ weeks). When it comes to turnover, Ledoit and Wolf (2003) report that the principal component estimator (our PCL estimator) involves less rebalancing than the LWI and LWF estimators. This is the case in our experiments too (see the column SPX of Table 7). Hence, notwithstanding the differences between the frameworks (different universe, sampling frequency for returns, number of factors, rebalancing period), our results are consistent with theirs.

Another result from the literature which is confirmed by our results is that weight constraints do not systematically reduce volatility, for any estimator and any sample size. Let us consider for example the SAM estimator and the smallest sample (fifty-two weekly returns). The long-only MV portfolio in universes Ind10, Ind17 and Ind48 (see Table 8) has lower volatility than its long-short counterpart (see Table 6). This is especially true in the universe Ind48, where the long-short portfolio has a very high volatility (37.3%, versus 12.3% for the long-only version). However, in the datasets FF6 and FF25, the opposite effect is observed: the long-only portfolios are more volatile. For other estimators, the impact of short-sales constraints is even less clear: it can be negligible (e.g., PCR estimator with Ind48 and $T = 104$), reduce the volatility (e.g., FCW estimator with FF6 and $T = 104$) or increase it (e.g., SAM estimator with FF6 and $T = 104$). Jagannathan and Ma (2003) report similar effects: imposing short-sales constraints can reduce the out-of-sample volatility of estimated MV portfolios based on the sample estimator in large datasets, bringing it below that of the naively diversified equally-weighted portfolio, but if a factor-based or shrunk estimator is used, the constraints lead instead to a higher volatility.

A third result already present in the literature is that optimal long-only portfolios are less sensitive than long-short ones to the choice of a regularization method. This is apparent from the results of Jagannathan and Ma (2003), and was also shown theoretically by Fan et al. (2012) and empirically by Pantaleo et al. (2011). As a matter of fact, volatilities across estimation methods for given sample size and dataset are closer to each other in Table 8 (long-only portfolios) than in Table 6. Moreover, volatilities in Table 8 display less sensitivity to the sample size than those in Table 6, as can be seen by comparing the various panels for a given dataset and a given estimation method. For the smallest datasets (i.e., those with less than 96 assets), volatility differences across estimators other than the sample one are hardly noticeable. A possible explanation is that the computation of optimal long-only weights does not require the inversion of covariance matrix estimator. The inversion process is known to be an error propagator (see for instance the bounds in Fan et al. (2008) for the sample covariance matrix) and therefore bypassing it can reduce estimation errors.

5.2.2. *Best and worst estimators for a given dataset and a given sample size.* Beside these already known results, Tables 6, 8 and 10 contain new information on the ability of various estimators to reduce the out-of-sample volatility of estimated MV portfolios. When short positions are allowed (Table 6), the LW estimators perform well: for each sample size and each dataset but FF96, the minimum volatility is achieved by one of these. It is only in the FF96 dataset that the minimum volatility is attained by a PC estimator, but in that case, the distance of the best LW estimator to the minimum does not exceed 40 basis points. The performances of PC and LW estimators, which are good, contrast with those of the benchmark estimators: for each sample size and each dataset, the highest volatility corresponds to either the SAM, CC or ID estimator, and most of the time to the ID one. This result is encouraging, because it shows that it is easy to improve the volatility of the naively diversified equally-weighted portfolio through scientific diversification.

For optimal long-only portfolios (Table 8), the LW estimators again show up among the best performers. As noted previously, their dominance is not very significant in the smallest datasets (that is, all datasets except FF96 and SPX), and in these universes, the PC estimators and the FEW, FCW and FFF ones give very close results. In the FF96 and SPX universes, the minimum is always achieved by a LW estimator, except with the FF96 dataset and $T = 208$: with this combination, the PCR estimator yields the lowest volatility (13.9%), which is only slightly below that obtained with the LWI estimator (14.0%). But the PC estimators do not lead to substantially different volatilities: for each sample size, the worst PC estimator gives a volatility that does not exceed the minimum by more than 40 basis points. For portfolios computed by rescaling, an unexpected result occurs: the minimum volatility for given dataset and sample size is often attained by the CC estimator. Even when this is not the case, as in the panel corresponding to $T = 208$ for the datasets Ind48, FF6, FF25 and FF96, the volatility of the MV portfolio based on the CC estimator is not larger than the minimum by more than 20 basis points. Thus, this estimator shows surprisingly good results when it is combined with the rescaling procedure. The PC estimators, however, are not much worse: for instance, the volatility of the portfolio constructed with the PCR estimator is always greater than or equal to the minimum, but the distance to the minimum does not exceed 50 basis points. The good performance of the CC estimator in this table contrasts with its poor results in the Monte-Carlo study (see Section 4).

It is also interesting to compare the various factor selection criteria, which correspond respectively to the PCL, PCP and PCR estimators. When short positions are allowed (Table 6), the PC estimator that yields the lowest volatility is in general the PCL or the PCP one, except for a few combinations of dataset and sample size. When short positions are ruled out in an optimal way (Table 8) and the sample counts 52 or 104 weeks, the PCL and PCP estimators also dominate the PCR one in most datasets. Moving to a large sample with 208 observations, we observe that the PCR estimator is the best among PC estimators for almost all datasets. Finally, when weights are obtained through rescaling (Table 10), it is the PCR estimator that dominates the PCL and PCP ones for the vast majority of datasets and sample sizes.

5.2.3. *Turnovers.* Turnovers are reported in Tables 7, 9 and 11. As expected, they are huge for long-short portfolios, in Table 7, and the numbers look all the more impressive if one recalls that these are quarterly turnovers. The SAM estimator, in particular, gives completely unreasonable turnovers, that often exceed 100%. It is clearly the ID estimator that implies the most acceptable levels, because the MV portfolio associated with this estimator is in fact the equally-weighted portfolio. The PC, LW, FEW, FCW and FFF estimators have comparable turnovers, which in most cases are lower than that of the MV portfolio based on the SAM estimator, but the levels are still very high. More reasonable figures, most of which are below 100%, are obtained for optimal long-only portfolios (Table 9) and for long-only portfolios with rescaled weights (Table 11). Again, the highest values correspond to the SAM estimator for most combinations of dataset and sample size.

The ID estimator set apart, if the weights are unconstrained, then the CC estimator ranks first in most cases and has the overall smallest turnover on average, closely followed by the FCW and FEW estimators, which is in line with the results of Ledoit and Wolf (2003). In contrast, the PCP, LWF, LWI and sample estimator often perform poorly. The shrinkage and sample estimators display particularly disappointing results in large datasets (Ind48, FF96 and SPX), while the last and second-to-last estimators in small datasets (FF6, Ind10 and Ind17) are the PCP and the sample estimators. In the case of long-only portfolios, whether computed in an optimal way or through rescaling (see Tables 9 and 11), the two structured estimators (CC and FEW) still display low turnovers. The PCR estimator implies reasonable levels of turnover for all datasets and all sample sizes, and for all datasets but the FF6 universe, the associated turnovers are lower than those of the shrunk estimators.

5.2.4. *Impact of rescaling.* One may also compare the volatilities of portfolios obtained by constraining weights to be nonnegative (Table 8) to those of portfolios with rescaled weights (Table 10). Unlike in the Monte-Carlo study, where it does not always lead to higher volatility than the optimization under nonnegativity constraints, the rescaling procedure here increases the volatility. This comes as no surprise given that the corresponding portfolios do not ex-ante minimize volatility. As far as turnover is concerned, Table 11 shows that for each dataset and sample size, the spread between the maximum and the minimum turnovers tends to be higher for rescaled weights than for long-only optimal weights, as was the case for volatilities.

6. Conclusion

6.1. **Summary of results.** The conclusions of the Monte-Carlo and empirical studies do not designate an estimator that would dominate the others consistently across all universes and all sample sizes. In order to have a synthetic view of the results, we score estimators after the figures reported in Tables 4, 6, 10, 7 and 11 (Table 5 is not used here because it serves only for outlier detection, and Tables 8 and 9 are not taken into account because they do not show enough significant differences between estimators). For each table, we group cells into categories: Gaussian versus Student returns for the simulated data, unconstrained versus rescaled weights for real data, and "low" or "high" ratio N/T for both. The performance indicator of each estimator is then averaged across all the cells that belong to a category. This indicator is the relative increase in volatility (RIV) for simulated data, and the out-of-sample volatility and the turnover for real data. We thus obtain a total of ten average criteria for the Monte-Carlo study, and twelve for the study on real data. An estimator is assigned a "+" in Table 12 if it ranks among those that lead to the lowest four volatilities or turnovers, and a "$-$" if it belongs to those that imply the highest four volatilities or turnovers. This ranking of estimators shows that the LWI estimator generates low out-of-sample volatilities when weights are unconstrained, but does it at the expense of a high turnover. It performs less well in terms of volatility when weights are rescaled. The LWF estimator displays a similar behavior: it demonstrates good ability to reduce the volatility in the real dataset when weights are unconstrained, but not when they are rescaled. Moreover, the MV portfolio built from this estimator is one of those with the highest turnovers.

The opposite occurs for the CC and FEW estimators: in the real dataset, they are better ranked in terms of volatility when short sales are allowed for than when weights are rescaled. The corresponding MV portfolios also rank among those that have the lowest turnovers. However, they do not have convincing performances in the Monte-Carlo study: the CC estimator is even one of those that yield the highest four RIVs.

The PC estimators tend to be good in terms of volatility reduction, both in the Monte-Carlo and in the empirical study. As the LW estimators, they are not well ranked in terms of turnover, but the LWF and LWI estimators have even worse rankings, as indicated by the many "$-$" grades that they obtain. Finally, only the PCL and the PCR estimators do not record any $-$ sign, which shows that they are never among the "worst" estimators according to the criteria retained in this table.

6.2. **Concluding remarks.** In this paper, we conduct an extensive comparison of twelve estimators of the covariance matrix, which can be categorized into three benchmark estimators (the sample estimator, the naive estimator equal to the identity matrix and the constant-correlation estimator), and a set of "regularized" estimators that encompasses a family of estimators based on an implicit factor model, a family of shrunk estimators and a family of estimators based on an explicit factor model. We run a Monte-Carlo study in order to compare the estimators across many paths for asset returns, and we consider two distributions for returns, with thin or fat tails. We next do an out-of-sample study on real data, thereby following the usual protocol in the empirical literature. In each of these parts, we consider different universes and different sample sizes, because these are two important drivers of estimation risk. We also repeat the analysis for long-short minimum variance portfolios, and for long-only portfolios with optimal weights or with weights obtained through a heuristic rescaling rule. Estimators are compared on the grounds of the volatility of the estimated minimum variance portfolio, which is the natural and standard criterion, but we also compute the turnovers, which are important for practitioners.

The results that we obtain do not demonstrate an unambiguous superiority of one regularized estimator of the covariance matrix over another one. They do confirm that the sample estimator and the identity matrix (which leads to proxying the minimum variance portfolio as the equally-weighted portfolio) are in general dominated in terms of volatility by at least one regularized estimator. But that is not to say that regularized estimators always perform well: a minimum variance portfolio constructed with one of them can have higher volatility than a similar portfolio using a benchmark estimator. The benchmark estimator that proves to be the most serious competitor for regularized estimators is the constant-correlation one. But it has very disappointing performance in the Monte-Carlo study when returns follow a Student distribution, which does not lead to consider it a valuable choice.

In order to choose an estimator, one may look at averages of minimum variance portfolio volatilities and turnovers across the various combinations of datasets, sample sizes and weight constraints that we have considered. According to this procedure, the PCL (Principal Component selection with Logarithmic criterion) and the PCR estimators (Principal Component selection with Random Matrix Theory) appear as good choices, because they perform well in terms of volatility reduction without implying too large levels of turnover.

Appendix A. Formulas for the shrinkage intensities

We denote $R_t^{(k)}$ the return of asset k at time t for $t = 1, \ldots, T$ and $\bar{R}^{(k)} = T^{-1} \sum_{t=1}^T R_t^{(k)}$ the average return over the estimation window. The bold notation $\boldsymbol{R}_t$ will stand for the vector of all asset returns at time t and $\bar{\boldsymbol{R}} = T^{-1} \sum_{t=1}^T \boldsymbol{R}_t$ for the vector of the averages of the assets' returns. The sample covariance matrix is then equal to $\widehat{\boldsymbol{\Sigma}} = (T-1)^{-1} \sum_{t=1}^T (\boldsymbol{R}_t - \bar{\boldsymbol{R}})(\boldsymbol{R}_t - \bar{\boldsymbol{R}})'$; its elements will be denoted $\hat{\sigma}_{k,l}$. The following constant will be used in two estimators of $\hat{\alpha}^*$:

$$\hat{\pi} = \sum_{k=1}^N \sum_{l=1}^N \hat{\pi}_{k,l},$$

where

$$\hat{\pi}_{k,l} = \frac{1}{T} \sum_{t=1}^T ((R_t^{(k)} - \bar{R}_k)(R_t^{(l)} - \bar{R}_j) - \hat{\sigma}_{k,l})^2.$$

A.1. **Shrinkage towards the single-factor covariance matrix.** In this case, the shrinkage target is $\widehat{\boldsymbol{\Sigma}}^F$ (defined in (9)) and we write $\hat{\sigma}_{k,l}^F$ for its elements. The returns of the market factor are denoted $R_t^{(0)}$ and their mean $R^{(0)}$. We write $\hat{\sigma}_{k,0}$ for the sample covariance between the returns of asset k and those of the market factor. In addition to $\hat{\pi}$, we require two other constants:

$$\hat{\gamma} = \sum_{k=1}^N \sum_{l=1}^N (\hat{\sigma}_{k,l} - \hat{\sigma}_{k,l}^F)^2$$

and

$$\hat{\rho} = \sum_{k=1}^N \pi_{k,k} + \sum_{k=1}^N \sum_{\substack{1 \le l \le N \\ l \ne k}} \sum_{t=1}^T \frac{q_{k,l,t}}{T}$$

where

$$\begin{aligned} q_{k,l,t} = &\frac{\hat{\sigma}_{l,0}\hat{\sigma}_{0,0}(R_t^{(k)} - \bar{R}^{(k)} + \hat{\sigma}_{k,0}\hat{\sigma}_{0,0}(R_t^{(l)} - \bar{R}^{(l)}) - \hat{\sigma}_{k,0}\hat{\sigma}_{l,0}(R_t^{(0)} - \bar{R}^{(0)})}{\hat{\sigma}_{0,0}^2}(R_t^{(0)} - \bar{R}^{(0)})(R_t^{(k)} - \bar{R}^{(k)})(R_t^{(l)} - \bar{R}^{(l)}) \\ &- \hat{\sigma}_{k,l}^F \hat{\sigma}_{k,l}^S, \end{aligned}$$

Ledoit and Wolf (2003) show that an estimator for the optimal shrinkage intensity is given by

$$\hat{\alpha}^* = \max\left(0, \min\left(1, \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma} T}\right)\right). \tag{22}$$

A.2. **Shrinkage towards the constant-correlation covariance matrix.** Ledoit and Wolf (2004a) provide the following estimator for the optimal shrinkage intensity when the shrinkage target is $\widehat{\boldsymbol{\Sigma}}^{CC}$ given in (6). We denote the element of this matrix $\hat{\sigma}_{k,l}^{CC}$. We start by introducing some intermediate constants. Given

$$\hat{\theta}_{kk,kl} = \frac{1}{T} \sum_{t=1}^T ((R_t^{(k)} - \bar{R}^{(k)})^2 - \hat{\sigma}_{k,k})((R_t^{(k)} - \bar{R}^{(k)})(R_t^{(l)} - \bar{R}^{(l)}) - \hat{\sigma}_{k,l}),$$

$$\hat{\theta}_{ll,kl} = \frac{1}{T} \sum_{t=1}^T ((R_t^{(l)} - \bar{R}^{(l)})^2 - \hat{\sigma}_{l,l})((R_t^{(k)} - \bar{R}^{(k)})(R_t^{(l)} - \bar{R}^{(l)}) - \hat{\sigma}_{k,l}),$$

we compute

$$\hat{\rho} = \sum_{k=1}^{N} \pi_{k,k} + \sum_{k=1}^{N} \sum_{\substack{1 \leq l \leq N \\ l \neq k}} \frac{\bar{r}}{2}, \left(\sqrt{\frac{\hat{\sigma}_{ll}}{\hat{\sigma}_{k,k}}} \hat{\theta}_{kk,kl} + \sqrt{\frac{\hat{\sigma}_{k,k}}{\hat{\sigma}_{l,l}}} \hat{\theta}_{ll,kl} \right),$$

$$\hat{\gamma} = \sum_{k=1}^{N} \sum_{l=1}^{N} (\hat{\sigma}_{k,l} - \hat{\sigma}_{k,l}^{CC})^2,$$

where $\hat{r}$ is given in Equation (5). Ledoit and Wolf (2004a) show that an estimator for $\hat{\alpha}^*$ is again given by (22).

A.3. **Shrinkage towards the scaled identity matrix.** Ledoit and Wolf (2004b) provide the following estimator for the optimal shrinkage intensity when the shrinkage target is $\bar{\sigma}^2 \boldsymbol{I}_N$ where $\bar{\sigma}^2 = \text{tr}(\widehat{\boldsymbol{\Sigma}})/N$. Given

$$d = \text{tr}((\widehat{\boldsymbol{\Sigma}} - \bar{\sigma}^2 \boldsymbol{I}_N)^2)/N$$

and

$$b = \min \left(d, \frac{1}{NT^2} \sum_{k=1}^{T} \text{tr}((\boldsymbol{R}_t - \bar{\boldsymbol{R}})(\boldsymbol{R}_t - \bar{\boldsymbol{R}})' - \widehat{\boldsymbol{\Sigma}}) \right),$$

the estimator of the optimal intensity is

$$\hat{\alpha}^* = b/d.$$

Appendix B. Tables

Table 2. List of estimators considered in this paper.

Name	**Abbrev.**	**Reference**	**Definition**
Principal component Logarithm	PCL	-	3.9
Principal component Penalized	PCP	Bai and Ng (2002)	3.7
Principal component RMT	PCR	Laloux et al. (2000)	3.8
Ledoit & Wolf constant correlation	LWC	Ledoit and Wolf (2004a)	3.11
Ledoit & Wolf single factor	LWF	Ledoit and Wolf (2003)	3.10
Ledoit & Wolf identity	LWI	Ledoit and Wolf (2004b)	3.12
Sample	SAM	-	3.1
Constant correlation	CC	Elton and Gruber (1973)	3.2
Identity	ID	-	3.3
Single factor equal weighted	FEW	-	3.4
Single factor cap weighted	FCW	-	3.5
Three factors Fama-French	FFF	-	3.6

This table contains the list of all covariance matrix estimators considered in this paper, with their abbreviated names. When available, a reference paper is cited. The last column refers to the formal definition of each estimator in the text.

TABLE 3. List of datasets considered in the section on real data.

Dataset	**Abbreviation**	**Number of assets**	**Time period**	**Source**
Ten industry portfolios representing the U.S. stock market	Ind10	10	01/1959-12/2012	KF
Seventeen industry portfolios representing the U.S. stock market	Ind17	17	01/1959-12/2012	KF
Forty-eight industry portfolios representing the U.S. stock market	Ind48	48	07/1969-12/2012	KF
Six Fama and French (1992) portfolios of firms sorted by size and book-to-market	FF6	6	01/1959-12/2012	KF
Twenty-five Fama and French (1992) portfolios of firms sorted by size and book-to-market	FF25	25	01/1959-12/2012	KF
Ninety-six Fama and French (1992) portfolios of firms sorted by size and book-to-market	FF96	96	07/1969-12/2012	KF
An adjusted (*) subset of the S&P 500 universe	SPX	$\in$(450,500)	04/1957-12/2010	CRSP

This table contains the list of the datasets on which the various estimators of the covariance matrix are compared, together with their abbreviation, their period of availability and the source of the data. "KF" means Kenneth French's website. The last dataset is a subset of the S&P 500 universe defined as follows: at each rebalancing date, we check if a stock belongs to the index. If it does, we check that past and future values are coherent for estimation and holding period purposes. If some data is missing or corrupt, then the asset is subtracted from the set. For our dataset and $T = 104$, this procedure gives 492 assets on average, with a minimum of 471 and a maximum of 500. Likewise, the FF96 was extracted from the FF100 returns downloaded from Kenneth French's website.

Table 4. Comparison of estimators on simulated data – Average of relative increase in volatility across estimators and true covariance matrices.

	Case	Dist.	N	T	PCL	PCP	PCR	LWC	LWF	LWI	SAM	CC	ID	FEW
	1	N	50	26	0.177	0.158	0.164	0.170	0.183	0.184	-	0.175		0.177
PANEL A:	2	N	50	52	0.137	0.122	0.129	0.119	0.136	0.149	4.661	0.130	0.256	0.134
LONG-SHORT	3	N	50	104	0.101	0.090	0.095	0.090	0.094	0.112	0.384	0.108		0.095
	4	N	50	208	0.071	0.065	0.064	0.067	0.061	0.075	0.145	0.098		0.063
	5	N	100	26	0.263	0.252	0.256	0.287	0.258	0.307	-	0.299		0.251
	6	N	100	52	0.193	0.187	0.189	0.216	0.188	0.250	-	0.239	0.446	0.186
	7	N	100	104	0.134	0.131	0.130	0.168	0.129	0.192	4.393	0.208		0.131
	8	N	100	208	0.090	0.088	0.085	0.126	0.084	0.136	0.387	0.192		0.090
	9	S	50	26	0.301	0.319	0.299	0.404	0.325	0.236	-	0.415		0.328
	10	S	50	52	0.284	0.288	0.286	0.349	0.295	0.227	5.466	0.363	0.256	0.306
	11	S	50	104	0.268	0.267	0.272	0.315	0.273	0.219	0.670	0.329		0.284
	12	S	50	208	0.248	0.247	0.251	0.288	0.248	0.207	0.372	0.306		0.259
	13	S	100	26	0.431	0.432	0.436	0.624	0.436	0.384	-	0.650		0.447
	14	S	100	52	0.397	0.396	0.397	0.564	0.388	0.369	-	0.593	0.446	0.413
	15	S	100	104	0.359	0.358	0.351	0.519	0.353	0.350	5.312	0.550		0.376
	16	S	100	208	0.318	0.318	0.304	0.479	0.319	0.324	0.710	0.521		0.339
	PANEL A Average				0.236	0.232	0.232	0.299	0.236	0.233	21.739	0.324	0.351	0.242
	17	N	50	26	0.168	0.147	0.152	0.147	0.170	0.163	-	0.152		0.166
PANEL B:	18	N	50	52	0.126	0.112	0.120	0.102	0.123	0.129	0.226	0.111	0.256	0.122
LONG ONLY	19	N	50	104	0.086	0.078	0.086	0.076	0.082	0.094	0.148	0.091		0.081
	20	N	50	208	0.054	0.051	0.055	0.056	0.051	0.062	0.081	0.082		0.051
	21	N	100	26	0.222	0.205	0.210	0.221	0.216	0.243	0.488	0.231		0.210
	22	N	100	52	0.167	0.156	0.160	0.162	0.161	0.195	0.355	0.177	0.446	0.159
	23	N	100	104	0.115	0.108	0.111	0.124	0.109	0.143	0.213	0.151		0.109
	24	N	100	208	0.072	0.069	0.071	0.092	0.068	0.094	0.120	0.138		0.069
	25	S	50	26	0.274	0.294	0.271	0.357	0.291	0.211	0.511	0.372		0.297
	26	S	50	52	0.255	0.259	0.258	0.311	0.265	0.202	0.386	0.322	0.256	0.274
	27	S	50	104	0.237	0.237	0.242	0.280	0.242	0.194	0.295	0.293		0.250
	28	S	50	208	0.218	0.217	0.220	0.257	0.219	0.183	0.231	0.273		0.225
	29	S	100	26	0.347	0.350	0.351	0.491	0.346	0.301	0.570	0.516		0.359
	30	S	100	52	0.319	0.320	0.320	0.440	0.313	0.288	0.471	0.434	0.446	0.331
	31	S	100	104	0.290	0.290	0.288	0.406	0.287	0.274	0.359	0.431		0.301
	32	S	100	208	0.257	0.257	0.251	0.374	0.259	0.254	0.279	0.408		0.271
	PANEL B Average				0.200	0.197	0.198	0.244	0.200	0.189	0.322	0.261	0.351	0.205
	33	N	50	26	0.145	0.132	0.136	0.142	0.144	0.163	-	0.147		0.140
PANEL C:	34	N	50	52	0.107	0.098	0.103	0.102	0.103	0.133	0.352	0.109	0.256	0.101
RESCALED	35	N	50	104	0.074	0.069	0.073	0.077	0.070	0.100	0.160	0.091		0.069
	36	N	50	208	0.049	0.046	0.047	0.058	0.045	0.068	0.086	0.082		0.045
	37	N	100	26	0.208	0.198	0.200	0.210	0.200	0.249	-	0.218		0.195
	38	N	100	52	0.158	0.151	0.152	0.163	0.150	0.208	-	0.175	0.446	0.147
	39	N	100	104	0.114	0.109	0.108	0.132	0.106	0.164	0.386	0.154		0.105
	40	N	100	208	0.080	0.076	0.074	0.105	0.072	0.122	0.171	0.142		0.072
	41	S	50	26	0.257	0.258	0.260	0.343	0.270	0.210	-	0.354		0.279
	42	S	50	52	0.238	0.236	0.240	0.304	0.245	0.202	0.425	0.315	0.256	0.256
	43	S	50	104	0.220	0.218	0.220	0.276	0.225	0.194	0.275	0.288		0.233
	44	S	50	208	0.200	0.201	0.196	0.254	0.204	0.184	0.215	0.271		0.211
	45	S	100	26	0.339	0.337	0.345	0.456	0.337	0.302	-	0.477		0.353
	46	S	100	52	0.311	0.310	0.312	0.422	0.305	0.290	-	0.442	0.446	0.324
	47	S	100	104	0.283	0.281	0.274	0.396	0.279	0.277	0.461	0.420		0.300
	48	S	100	208	0.249	0.250	0.234	0.371	0.252	0.257	0.290	0.402		0.266
	PANEL C Average				0.190	0.186	0.186	0.238	0.188	0.195	-	0.255	0.351	0.194
	Gaussian Dist. Average				0.130	0.121	0.124	0.134	0.125	0.156	6.157	0.154	0.351	0.124
	Student Dist. Average				0.288	0.289	0.287	0.387	0.291	0.256	8.791	0.406	0.351	0.303
	Average for $N/T < 0.5$				0.161	0.158	0.158	0.202	0.159	0.160	0.279	0.229	0.319	0.165
	Average for $N/T > 1.5$				0.260	0.256	0.256	0.323	0.260	0.253	-	0.337	0.383	0.265
	Overall Average				0.209	0.205	0.205	0.260	0.208	0.206	-	0.280	0.351	0.214

For each estimation method, each universe size, each sample size (expressed in weeks) and each distribution of returns (normal or Student), we compute three estimated MV portfolios: a long-short one (Panel A), a long-only optimal one (Panel B), and a long-only one in which negative weights have been eliminated by a heuristic adjustment to the long-short portfolio (Panel C). The table shows the average relative increases in volatility (RIV) with respect to the MV portfolios that would use the true covariance matrix. The RIV for a given true matrix and a given estimator are defined in (17) for Panel A, in (20) for Panel B and in (21) for Panel C. Averages are then taken across 50 estimators and 527 true covariance matrices. The estimators are defined in Table 2. Missing figures for the SAM estimator arise when $N > T$ and the sample covariance matrix is not invertible.

Table 5. Comparison of estimators on simulated data – Average (across true covariance matrices) of standard deviations (across estimators) of relative increases in volatility.

	Case	Dist.	N	T	PCL	PCP	PCR	LWC	LWF	LWI	SAM	CC	ID	FEW
	1	N	50	26	0.043	0.036	0.039	0.040	0.043	0.031	-	0.040		0.042
PANEL A:	2	N	50	52	0.032	0.028	0.031	0.022	0.030	0.027	3.517	0.022	0	0.030
LONG-SHORT	3	N	50	104	0.023	0.021	0.021	0.016	0.019	0.021	0.093	0.015		0.020
	4	N	50	208	0.016	0.014	0.013	0.012	0.012	0.015	0.031	0.010		0.012
	5	N	100	26	0.044	0.042	0.044	0.046	0.044	0.043	-	0.046		0.043
	6	N	100	52	0.035	0.034	0.034	0.028	0.033	0.037	-	0.028	0	0.033
	7	N	100	104	0.024	0.024	0.023	0.021	0.022	0.028	2.158	0.019		0.023
	8	N	100	208	0.015	0.017	0.014	0.017	0.014	0.020	0.065	0.014		0.015
	9	S	50	26	0.062	0.078	0.061	0.120	0.093	0.029	-	0.119		0.069
	10	S	50	52	0.059	0.063	0.061	0.087	0.063	0.025	4.331	0.089	0	0.064
	11	S	50	104	0.058	0.059	0.061	0.071	0.058	0.026	0.153	0.068		0.062
	12	S	50	208	0.056	0.055	0.058	0.061	0.056	0.028	0.080	0.055		0.061
	13	S	100	26	0.056	0.058	0.054	0.127	0.120	0.038	-	0.125		0.061
	14	S	100	52	0.058	0.057	0.058	0.101	0.076	0.038	-	0.097	0	0.063
	15	S	100	104	0.058	0.056	0.059	0.087	0.061	0.044	2.515	0.080		0.066
	16	S	100	208	0.058	0.055	0.053	0.081	0.060	0.052	0.113	0.067		0.069
	17	N	50	26	0.044	0.037	0.039	0.038	0.043	0.030	0.093	0.036		0.042
PANEL B:	18	N	50	52	0.032	0.028	0.031	0.021	0.029	0.025	0.058	0.020	0	0.029
LONG ONLY	19	N	50	104	0.021	0.019	0.021	0.015	0.019	0.020	0.033	0.013		0.019
	20	N	50	208	0.013	0.012	0.012	0.011	0.011	0.013	0.018	0.009		0.011
	21	N	100	26	0.041	0.037	0.039	0.041	0.041	0.035	0.121	0.040		0.040
	22	N	100	52	0.033	0.030	0.032	0.025	0.031	0.031	0.065	0.024	0	0.031
	23	N	100	104	0.023	0.021	0.022	0.018	0.021	0.025	0.039	0.017		0.021
	24	N	100	208	0.014	0.014	0.014	0.014	0.013	0.017	0.022	0.012		0.013
	25	S	50	26	0.063	0.082	0.061	0.108	0.069	0.028	0.122	0.109		0.069
	26	S	50	52	0.058	0.064	0.061	0.079	0.061	0.024	0.092	0.076	0	0.063
	27	S	50	104	0.055	0.056	0.059	0.062	0.056	0.024	0.069	0.059		0.059
	28	S	50	208	0.054	0.053	0.055	0.054	0.054	0.027	0.055	0.047		0.057
	29	S	100	26	0.054	0.057	0.052	0.111	0.058	0.033	0.140	0.109		0.057
	30	S	100	52	0.053	0.054	0.055	0.081	0.054	0.033	0.091	0.078	0	0.057
	31	S	100	104	0.053	0.053	0.054	0.069	0.052	0.037	0.068	0.062		0.057
	32	S	100	208	0.052	0.050	0.050	0.061	0.052	0.042	0.052	0.051		0.059
	33	N	50	26	0.034	0.030	0.032	0.034	0.034	0.028	-	0.033		0.034
PANEL C:	34	N	50	52	0.026	0.023	0.026	0.020	0.025	0.024	0.070	0.019	0	0.024
RESCALED	35	N	50	104	0.018	0.016	0.018	0.014	0.016	0.019	0.035	0.013		0.016
	36	N	50	208	0.012	0.011	0.011	0.011	0.010	0.013	0.019	0.009		0.010
	37	N	100	26	0.035	0.032	0.033	0.033	0.033	0.032	-	0.034		0.032
	38	N	100	52	0.027	0.025	0.026	0.022	0.025	0.027	-	0.022	0	0.025
	39	N	100	104	0.019	0.018	0.019	0.017	0.018	0.022	0.052	0.016		0.018
	40	N	100	208	0.013	0.012	0.012	0.013	0.012	0.016	0.026	0.011		0.012
	41	S	50	26	0.050	0.053	0.050	0.089	0.058	0.027	-	0.090		0.059
	42	S	50	52	0.046	0.047	0.048	0.069	0.051	0.023	0.085	0.068	0	0.053
	43	S	50	104	0.045	0.044	0.047	0.056	0.047	0.023	0.057	0.053		0.050
	44	S	50	208	0.043	0.042	0.044	0.050	0.046	0.026	0.044	0.044		0.049
	45	S	100	26	0.044	0.044	0.044	0.078	0.049	0.030	-	0.079		0.049
	46	S	100	52	0.043	0.042	0.044	0.063	0.044	0.031	-	0.062	0	0.049
	47	S	100	104	0.043	0.041	0.042	0.056	0.043	0.034	0.063	0.054		0.049
	48	S	100	208	0.042	0.039	0.038	0.053	0.043	0.039	0.041	0.046		0.051

For each estimation method, each universe size, each sample size (expressed in weeks) and each distribution of returns (normal or Student), we compute three estimated MV portfolios: a long-short one (Panel A), a long-only optimal one (Panel B), and a long-only one in which negative weights have been eliminated by a heuristic adjustment to the long-short portfolio (Panel C). The table shows the average (across the 527 true covariance matrices) of the standard deviations (across the 50 estimators) of the relative increases in volatility (RIV). The RIV measures the increase in volatility which arises from the use of an imperfect estimate as opposed to the true covariance matrix. The RIV for a given true matrix and a given estimator are defined in (17) for Panel A, in (20) for Panel B and in (21) for Panel C. Averages are then taken across 50 estimators and 527 true covariance matrices. The estimators are defined in Table 2. Missing figures for the SAM estimator arise when $N > T$ and the sample covariance matrix is not invertible.

TABLE 6. Comparison of estimators on real data – Out-of-sample volatilities of estimated long-short minimum variance portfolios.

	Ind10	**Ind17**	**Ind48**	**FF6**	**FF25**	**FF96**	**SPX**	**Average**
Long-short				$T = 52$ weeks				
PCL	0.125	0.123	0.118	0.136	0.111	0.103	0.100	0.117
PCP	0.124	0.136	0.118	0.123	0.113	0.103	0.104	0.117
PCR	0.122	0.122	0.120	0.138	0.127	0.115	0.099	0.120
LWC	0.121	0.122	0.117	0.135	0.127	0.115	0.107	0.121
LWF	0.120	0.119	0.112	0.122	0.108	0.105	0.098	0.112
LWI	0.118	0.117	0.118	0.126	0.108	0.108	0.108	0.115
SAM	0.124	0.129	0.373	0.123	0.126	-	-	-
CC	0.126	0.130	0.133	0.136	0.134	0.121	0.129	0.130
ID	0.150	0.160	0.170	0.158	0.161	0.169	0.164	0.162
FEW	0.122	0.122	0.120	0.137	0.128	0.118	0.107	0.122
FCW	0.122	0.124	0.122	0.143	0.134	0.126	0.136	0.130
FFF	0.122	0.121	0.115	0.132	0.110	0.103	0.132	0.119
				$T = 104$ weeks				
PCL	0.124	0.121	0.118	0.140	0.114	0.103	0.099	0.117
PCP	0.120	0.129	0.118	0.125	0.117	0.103	0.106	0.117
PCR	0.121	0.121	0.119	0.141	0.135	0.122	0.097	0.122
LWC	0.118	0.117	0.115	0.134	0.123	0.114	0.106	0.118
LWF	0.118	0.116	0.113	0.125	0.110	0.107	0.097	0.112
LWI	0.117	0.115	0.116	0.128	0.109	0.111	0.107	0.115
SAM	0.120	0.119	0.132	0.125	0.114	0.301	-	-
CC	0.125	0.130	0.136	0.138	0.146	0.136	0.139	0.136
ID	0.150	0.160	0.170	0.158	0.161	0.169	0.164	0.162
FEW	0.121	0.122	0.124	0.141	0.139	0.131	0.112	0.127
FCW	0.122	0.124	0.126	0.146	0.143	0.137	0.140	0.134
FFF	0.121	0.121	0.117	0.133	0.114	0.107	0.136	0.121
				$T = 208$ weeks				
PCL	0.126	0.124	0.124	0.136	0.117	0.109	0.105	0.120
PCP	0.119	0.127	0.124	0.124	0.116	0.107	0.114	0.119
PCR	0.122	0.123	0.124	0.137	0.134	0.122	0.099	0.123
LWC	0.118	0.116	0.116	0.128	0.116	0.111	0.107	0.116
LWF	0.119	0.116	0.115	0.123	0.110	0.109	0.099	0.113
LWI	0.118	0.116	0.116	0.126	0.110	0.111	0.110	0.115
SAM	0.119	0.118	0.121	0.124	0.111	0.123	-	-
CC	0.126	0.132	0.141	0.135	0.141	0.137	0.151	0.138
ID	0.150	0.160	0.170	0.158	0.161	0.169	0.164	0.162
FEW	0.122	0.123	0.128	0.137	0.138	0.134	0.123	0.129
FCW	0.123	0.125	0.129	0.145	0.145	0.142	0.143	0.136
FFF	0.121	0.121	0.122	0.130	0.117	0.114	0.141	0.124

For each dataset (see definitions and time spans in Table 3), each sample size (as measured by T) and each of the twelve regularization methods, we estimate one covariance matrix per quarter and we compute estimated minimum variance weights without imposing weight constraints. The table reports the out-of-sample volatility of these portfolios. The estimators are defined in Table 2. Missing figures for the SAM estimator arise when $N > T$, because the sample covariance matrix is not invertible.

Table 7. Comparison of estimators on real data – Turnovers of estimated long-short minimum variance portfolios.

	Ind10	**Ind17**	**Ind48**	**FF6**	**FF25**	**FF96**	**SPX**	**Average**
Long-short					$T = 52$			
PCL	1.062	1.159	1.431	1.012	2.413	2.485	1.194	1.537
PCP	1.440	2.588	1.683	1.771	3.631	2.384	0.985	2.069
PCR	0.820	1.017	1.217	1.045	1.603	1.676	1.218	1.228
LWC	0.740	1.002	1.873	0.858	1.572	2.041	1.640	1.389
LWF	1.071	1.382	1.755	1.555	3.210	3.439	1.329	1.963
LWI	0.749	1.153	2.097	0.703	2.028	3.791	1.368	1.698
SAM	1.440	2.421	21.386	1.771	6.482	-	-	-
CC	0.644	0.813	0.926	0.855	1.391	1.590	0.997	1.031
ID	0.043	0.047	0.061	0.031	0.035	0.044	0.106	0.052
FEW	0.783	0.969	1.118	1.024	1.530	1.590	0.921	1.134
FCW	0.779	0.964	1.090	0.993	1.377	1.483	0.374	1.009
FFF	0.952	1.160	1.394	2.197	2.241	2.264	0.566	1.539
					$T = 104$			
PCL	0.668	0.692	0.977	0.650	1.535	1.824	0.967	1.045
PCP	0.763	1.598	1.173	1.049	2.540	1.773	0.790	1.384
PCR	0.480	0.638	0.849	0.664	1.105	1.241	1.042	0.860
LWC	0.474	0.657	1.400	0.567	1.184	1.678	1.430	1.056
LWF	0.658	0.872	1.291	1.011	2.136	2.958	1.140	1.438
LWI	0.507	0.773	1.537	0.527	1.449	3.507	1.494	1.399
SAM	0.763	1.152	3.260	1.049	2.894	32.366	-	-
CC	0.379	0.507	0.609	0.544	0.959	1.106	0.682	0.684
ID	0.043	0.047	0.061	0.031	0.035	0.044	0.106	0.052
FEW	0.466	0.598	0.732	0.653	1.052	1.120	0.644	0.752
FCW	0.460	0.594	0.716	0.624	0.910	1.026	0.251	0.654
FFF	0.549	0.703	0.910	1.360	1.419	1.563	0.382	0.984
					$T = 208$			
PCL	0.406	0.444	0.681	0.407	0.922	1.243	0.828	0.704
PCP	0.409	0.989	0.898	0.597	1.582	1.276	0.666	0.917
PCR	0.293	0.422	0.621	0.408	0.735	1.018	0.910	0.630
LWC	0.299	0.429	0.919	0.358	0.802	1.347	1.320	0.782
LWF	0.381	0.540	0.875	0.588	1.283	2.185	1.040	0.985
LWI	0.316	0.485	0.953	0.352	0.939	2.346	1.781	1.025
SAM	0.409	0.623	1.355	0.597	1.475	4.169	-	-
CC	0.235	0.331	0.420	0.332	0.657	0.766	0.543	0.469
ID	0.043	0.047	0.061	0.031	0.035	0.044	0.106	0.052
FEW	0.285	0.384	0.506	0.405	0.697	0.775	0.482	0.505
FCW	0.284	0.390	0.497	0.389	0.607	0.705	0.185	0.437
FFF	0.323	0.432	0.602	0.773	0.879	1.047	0.252	0.615

For each dataset (see definitions and time spans in Table 3), each sample size (as measured by T) and each of the twelve regularization methods, we estimate one covariance matrix per quarter and we compute estimated minimum variance weights without imposing weight constraints. The table reports the turnover of these portfolios. The estimators are defined in Table 2. Missing figures for the SAM estimator arise when $N > T$, because the sample covariance matrix is not invertible.

TABLE 8. Comparison of estimators on real data – Out-of-sample volatilities of estimated minimum variance portfolios subject to long-only constraints.

	Ind10	**Ind17**	**Ind48**	**FF6**	**FF25**	**FF96**	**SPX**	**Average**
Optimal long-only				$T = 52$				
PCL	0.121	0.120	0.121	0.140	0.135	0.135	0.115	0.127
PCP	0.121	0.121	0.122	0.139	0.136	0.136	0.115	0.127
PCR	0.121	0.120	0.121	0.140	0.136	0.136	0.115	0.127
LWC	0.121	0.121	0.122	0.139	0.136	0.137	0.114	0.127
LWF	0.121	0.120	0.121	0.139	0.136	0.136	0.113	0.127
LWI	0.122	0.121	0.123	0.140	0.135	0.135	0.119	0.128
SAM	0.121	0.121	0.123	0.139	0.136	0.136	0.125	0.129
CC	0.122	0.123	0.123	0.140	0.136	0.138	0.116	0.128
ID	0.150	0.160	0.170	0.158	0.161	0.169	0.164	0.162
FEW	0.121	0.120	0.121	0.140	0.136	0.137	0.115	0.127
FCW	0.121	0.120	0.122	0.139	0.138	0.140	0.137	0.131
FFF	0.121	0.121	0.122	0.139	0.136	0.137	0.135	0.130
				$T = 104$				
PCL	0.121	0.120	0.121	0.141	0.137	0.139	0.112	0.127
PCP	0.121	0.121	0.121	0.139	0.138	0.139	0.113	0.127
PCR	0.121	0.120	0.119	0.141	0.138	0.139	0.113	0.127
LWC	0.121	0.120	0.120	0.140	0.138	0.141	0.111	0.127
LWF	0.121	0.120	0.120	0.139	0.137	0.139	0.111	0.127
LWI	0.122	0.121	0.121	0.140	0.137	0.138	0.114	0.128
SAM	0.121	0.120	0.121	0.139	0.137	0.140	0.117	0.128
CC	0.122	0.123	0.123	0.140	0.139	0.143	0.118	0.130
ID	0.150	0.160	0.170	0.158	0.161	0.169	0.164	0.162
FEW	0.121	0.120	0.120	0.141	0.139	0.142	0.115	0.128
FCW	0.121	0.121	0.121	0.140	0.139	0.143	0.140	0.132
FFF	0.121	0.121	0.121	0.139	0.138	0.139	0.138	0.131
				$T = 208$				
PCL	0.122	0.120	0.122	0.140	0.137	0.140	0.113	0.128
PCP	0.122	0.122	0.122	0.140	0.137	0.140	0.116	0.128
PCR	0.121	0.121	0.121	0.140	0.137	0.139	0.114	0.128
LWC	0.122	0.121	0.122	0.140	0.137	0.141	0.112	0.128
LWF	0.122	0.121	0.122	0.140	0.137	0.141	0.112	0.128
LWI	0.122	0.122	0.123	0.140	0.137	0.140	0.114	0.128
SAM	0.122	0.121	0.123	0.140	0.137	0.141	0.114	0.128
CC	0.123	0.122	0.126	0.140	0.140	0.144	0.124	0.131
ID	0.150	0.160	0.170	0.158	0.161	0.169	0.164	0.162
FEW	0.121	0.121	0.121	0.140	0.139	0.143	0.120	0.129
FCW	0.121	0.121	0.122	0.140	0.140	0.144	0.144	0.133
FFF	0.121	0.121	0.121	0.140	0.137	0.140	0.142	0.132

For each dataset (see definitions and time spans in Table 3), each sample size (as measured by T) and each of the twelve regularization methods, we estimate one covariance matrix per quarter and we compute estimated minimum variance weights by numerically solving Program (18). The table reports the out-of-sample volatility of these portfolios. The estimators are defined in Table 2. Missing figures for the SAM estimator arise when $N > T$, because the sample covariance matrix is not invertible.

Table 9. Comparison of estimators on real data – Turnovers of estimated minimum variance portfolios subject to long-only constraints.

	Ind10	**Ind17**	**Ind48**	**FF6**	**FF25**	**FF96**	**SPX**	**Average**
Optimal long-only	$T = 52$							
PCL	0.359	0.376	0.521	0.398	0.637	0.827	0.948	0.581
PCP	0.386	0.437	0.549	0.398	0.657	0.826	0.838	0.584
PCR	0.330	0.371	0.487	0.403	0.599	0.808	0.933	0.562
LWC	0.369	0.414	0.539	0.383	0.634	0.825	0.861	0.575
LWF	0.363	0.390	0.516	0.394	0.642	0.828	0.894	0.575
LWI	0.341	0.392	0.538	0.353	0.559	0.782	0.921	0.555
SAM	0.386	0.414	0.603	0.398	0.660	0.881	1.277	0.660
CC	0.369	0.412	0.486	0.385	0.622	0.802	0.734	0.544
ID	0.043	0.047	0.061	0.031	0.035	0.044	0.106	0.052
FEW	0.323	0.364	0.477	0.399	0.583	0.786	0.812	0.535
FCW	0.315	0.363	0.483	0.364	0.558	0.740	0.381	0.458
FFF	0.350	0.380	0.518	0.410	0.633	0.836	0.578	0.529
	$T = 104$							
PCL	0.207	0.212	0.304	0.243	0.412	0.543	0.617	0.363
PCP	0.216	0.267	0.320	0.232	0.428	0.544	0.530	0.362
PCR	0.183	0.218	0.298	0.248	0.388	0.538	0.625	0.357
LWC	0.213	0.227	0.337	0.223	0.412	0.561	0.545	0.360
LWF	0.209	0.230	0.315	0.231	0.414	0.555	0.573	0.361
LWI	0.207	0.231	0.332	0.225	0.369	0.525	0.647	0.362
SAM	0.216	0.237	0.348	0.232	0.424	0.585	0.815	0.408
CC	0.196	0.230	0.277	0.223	0.403	0.547	0.417	0.328
ID	0.043	0.047	0.061	0.031	0.035	0.044	0.106	0.052
FEW	0.180	0.210	0.279	0.245	0.378	0.517	0.484	0.328
FCW	0.174	0.206	0.288	0.218	0.343	0.474	0.254	0.280
FFF	0.189	0.214	0.306	0.231	0.419	0.535	0.383	0.325
	$T = 208$							
PCL	0.119	0.116	0.179	0.142	0.227	0.335	0.413	0.219
PCP	0.121	0.160	0.196	0.142	0.248	0.332	0.395	0.228
PCR	0.096	0.127	0.179	0.143	0.219	0.312	0.413	0.213
LWC	0.116	0.127	0.201	0.137	0.234	0.344	0.360	0.217
LWF	0.117	0.133	0.193	0.141	0.235	0.346	0.371	0.219
LWI	0.121	0.136	0.202	0.138	0.218	0.329	0.436	0.226
SAM	0.121	0.136	0.212	0.142	0.236	0.362	0.510	0.246
CC	0.111	0.125	0.158	0.137	0.240	0.336	0.248	0.194
ID	0.043	0.047	0.061	0.031	0.035	0.044	0.106	0.052
FEW	0.095	0.116	0.164	0.142	0.227	0.322	0.295	0.194
FCW	0.094	0.116	0.168	0.128	0.210	0.282	0.185	0.169
FFF	0.101	0.119	0.178	0.133	0.231	0.313	0.247	0.189

For each dataset (see definitions and time spans in Table 3), each sample size (as measured by T) and each of the twelve regularization methods, we estimate one covariance matrix per quarter and we compute estimated minimum variance weights by numerically solving Program (18). The table reports the turnover of these portfolios. The estimators are defined in Table 2. Missing figures for the SAM estimator arise when $N > T$, because the sample covariance matrix is not invertible.

TABLE 10. Comparison of estimators on real data – Out-of-sample volatilities of estimated minimum variance portfolios with rescaled weights.

	Ind10	**Ind17**	**Ind48**	**FF6**	**FF25**	**FF96**	**SPX**	**Average**
Rescaled				$T = 52$				
PCL	0.127	0.130	0.137	0.142	0.144	0.146	0.122	0.135
PCP	0.131	0.139	0.139	0.147	0.147	0.146	0.121	0.139
PCR	0.124	0.128	0.134	0.142	0.141	0.144	0.122	0.134
LWC	0.124	0.127	0.137	0.140	0.139	0.143	0.122	0.133
LWF	0.129	0.132	0.139	0.146	0.145	0.149	0.123	0.138
LWI	0.127	0.132	0.144	0.143	0.144	0.153	0.137	0.140
SAM	0.131	0.136	0.160	0.147	0.149	-	-	-
CC	0.124	0.127	0.133	0.140	0.139	0.142	0.118	0.132
ID	0.150	0.160	0.170	0.158	0.161	0.169	0.164	0.162
FEW	0.125	0.128	0.134	0.142	0.141	0.144	0.118	0.133
FCW	0.125	0.128	0.133	0.139	0.139	0.144	0.137	0.135
FFF	0.127	0.129	0.135	0.148	0.143	0.146	0.135	0.138
				$T = 104$				
PCL	0.127	0.130	0.138	0.143	0.145	0.148	0.128	0.137
PCP	0.130	0.138	0.140	0.147	0.149	0.148	0.121	0.139
PCR	0.124	0.128	0.134	0.142	0.143	0.146	0.124	0.134
LWC	0.124	0.127	0.138	0.141	0.141	0.146	0.125	0.135
LWF	0.128	0.132	0.140	0.147	0.147	0.151	0.126	0.139
LWI	0.127	0.132	0.144	0.144	0.145	0.154	0.139	0.141
SAM	0.130	0.134	0.150	0.147	0.148	0.163	0.150	-
CC	0.124	0.127	0.134	0.141	0.141	0.146	0.120	0.133
ID	0.150	0.160	0.170	0.158	0.161	0.169	0.164	0.162
FEW	0.124	0.129	0.135	0.143	0.143	0.147	0.119	0.134
FCW	0.125	0.129	0.134	0.140	0.141	0.147	0.140	0.137
FFF	0.126	0.130	0.135	0.149	0.145	0.148	0.138	0.139
				$T = 208$				
PCL	0.128	0.130	0.140	0.143	0.147	0.151	0.126	0.138
PCP	0.128	0.138	0.142	0.148	0.151	0.151	0.125	0.140
PCR	0.124	0.129	0.135	0.142	0.144	0.148	0.128	0.136
LWC	0.125	0.128	0.140	0.142	0.143	0.149	0.130	0.137
LWF	0.128	0.133	0.142	0.148	0.149	0.154	0.130	0.141
LWI	0.127	0.133	0.143	0.145	0.147	0.155	0.144	0.142
SAM	0.128	0.134	0.146	0.148	0.149	0.157	-	-
CC	0.124	0.127	0.136	0.142	0.143	0.149	0.123	0.135
ID	0.150	0.160	0.170	0.158	0.161	0.169	0.164	0.162
FEW	0.124	0.129	0.136	0.143	0.145	0.150	0.121	0.135
FCW	0.125	0.129	0.135	0.140	0.142	0.149	0.144	0.138
FFF	0.136	0.130	0.137	0.150	0.146	0.150	0.143	0.142

For each dataset (see definitions and time spans in Table 3), each sample size (as measured by T) and each of the twelve regularization methods, we estimate one covariance matrix per quarter and we compute estimated minimum variance weights as follows: first, unconstrained weights are computed, and second, the rescaling procedure described in (19) is applied to obtain nonnegative weights. The table reports the out-of-sample volatility of these portfolios. The estimators are defined in Table 2. Missing figures for the SAM estimator arise when $N > T$, because the sample covariance matrix is not invertible.

TABLE 11. Comparison of estimators on real data – Turnovers of estimated minimum variance portfolios with rescaled weights.

	Ind10	**Ind17**	**Ind48**	**FF6**	**FF25**	**FF96**	**SPX**	**Average**
Rescaled				$T = 52$				
PCL	0.359	0.360	0.431	0.286	0.428	0.500	0.509	0.410
PCP	0.454	0.609	0.483	0.395	0.530	0.486	0.450	0.487
PCR	0.308	0.340	0.385	0.319	0.370	0.407	0.516	0.378
LWC	0.341	0.382	0.516	0.300	0.403	0.506	0.562	0.430
LWF	0.388	0.430	0.487	0.378	0.532	0.616	0.544	0.482
LWI	0.344	0.417	0.557	0.273	0.450	0.634	0.529	0.458
SAM	0.454	0.554	1.197	0.395	0.704	-	-	-
CC	0.314	0.333	0.365	0.304	0.369	0.438	0.420	0.363
ID	0.043	0.047	0.061	0.031	0.035	0.044	0.106	0.052
FEW	0.298	0.329	0.362	0.305	0.354	0.398	0.427	0.353
FCW	0.298	0.321	0.355	0.299	0.347	0.387	0.349	0.337
FFF	0.344	0.364	0.416	0.385	0.408	0.474	0.477	0.410
				$T = 104$				
PCL	0.223	0.209	0.285	0.178	0.258	0.328	0.354	0.262
PCP	0.268	0.416	0.315	0.244	0.355	0.320	0.305	0.318
PCR	0.177	0.207	0.258	0.197	0.229	0.270	0.372	0.244
LWC	0.215	0.241	0.363	0.187	0.278	0.357	0.405	0.292
LWF	0.243	0.269	0.343	0.238	0.358	0.447	0.389	0.327
LWI	0.224	0.262	0.392	0.179	0.304	0.483	0.443	0.327
SAM	0.268	0.309	0.526	0.244	0.409	1.100	-	-
CC	0.177	0.199	0.226	0.189	0.228	0.273	0.289	0.226
ID	0.043	0.047	0.061	0.031	0.035	0.044	0.106	0.052
FEW	0.173	0.197	0.228	0.188	0.221	0.252	0.265	0.218
FCW	0.171	0.195	0.225	0.176	0.213	0.241	0.238	0.208
FFF	0.200	0.219	0.264	0.235	0.241	0.294	0.333	0.255
				$T = 208$				
PCL	0.134	0.125	0.183	0.108	0.143	0.197	0.267	0.165
PCP	0.156	0.267	0.228	0.140	0.213	0.202	0.225	0.204
PCR	0.102	0.129	0.172	0.116	0.131	0.189	0.276	0.159
LWC	0.128	0.152	0.232	0.121	0.167	0.243	0.311	0.193
LWF	0.147	0.168	0.226	0.139	0.204	0.305	0.299	0.213
LWI	0.139	0.165	0.243	0.111	0.180	0.315	0.397	0.221
SAM	0.156	0.179	0.277	0.140	0.218	0.411	-	-
CC	0.103	0.117	0.139	0.112	0.136	0.163	0.162	0.133
ID	0.043	0.047	0.061	0.031	0.035	0.044	0.106	0.052
FEW	0.101	0.118	0.144	0.112	0.129	0.154	0.179	0.134
FCW	0.100	0.119	0.144	0.102	0.124	0.145	0.178	0.130
FFF	0.115	0.134	0.165	0.128	0.137	0.176	0.229	0.155

For each dataset (see definitions and time spans in Table 3), each sample size (as measured by T) and each of the twelve regularization methods, we estimate one covariance matrix per quarter and we compute estimated minimum variance weights as follows: first, unconstrained weights are computed, and second, the rescaling procedure described in (19) is applied to obtain nonnegative weights. The table reports the turnover of these portfolios. The estimators are defined in Table 2. Missing figures for the SAM estimator arise when $N > T$, because the sample covariance matrix is not invertible.

Table 12. Comparison of estimators on simulated or real data – Summary of results.

Dataset	**Criterion**	**Condition**	**Table**	**PCL**	**PCP**	**PCR**	**LWC**	**LWF**	**LWI**	**SAM**	**CC**	**ID**	**FEW**	**FCW**	**FFF**
Simulated	RIV	Gaussian returns	4		+	+		+	-	-	-	-	+	NA	NA
Simulated	RIV	Student returns	4	+	+	+	-		+	-	-	-		NA	NA
Simulated	RIV	$N/T < 0.5$	4		+	+	-	+	+	-	-	-		NA	NA
Simulated	RIV	$N/T > 1.5$	4	+	+	+	-		+	-	-	-		NA	NA
Real	OOS Volatility	Unconstrained weights	6	+	+			+	+	-	-	-		-	
Real	OOS Volatility	Rescaled weights	10			+	+		-	-	+	-	+		-
Real	OOS Volatility	$N/T < 0.5$	6 & 10				+	+	+		-	-	-	-	+
Real	OOS Volatility	$N/T > 1.5$	6 & 10	+	+	+		+		-		-		-	-
Real	Turnover	Unconstrained weights	7		-			-	-	-	+	+	+	+	
Real	Turnover	Rescaled weights	11		-			-	-	-	+	+	+	+	
Real	Turnover	$N/T < 0.5$	7 & 11		-			-		-	+	+	+	+	-
Real	Turnover	$N/T > 1.5$	7 & 11					-	-	-		+	+	+	+

This table provides a comparison of the twelve covariance matrix estimators considered in this paper in various frameworks. For simulated data, a distinction is made between Gaussian and Student returns, as well as between a "low" ratio between universe size (N) and sample size (T), and a "high" ratio N/T. For real data, a distinction is made between unconstrained estimated minimum variance (MV) portfolios and long-only portfolios obtained through the heuristic rescaling procedure (19), as well as between low and high ratio N/T. The criterion used to compare estimators is the relative increase in volatility (RIV) associated with the estimated MV portfolio within the Monte-Carlo framework, and the out-of-sample volatility of the estimated portfolio within the real data framework. The criteria are averaged across all cells that satisfy the specified condition in the table specified in the fourth column. An estimator receives a "+" if its average criterion ranks among lowest four values, and a "−" if it ranks among the highest four. The sign "NA" ("non available") means that the estimator was not computed. The estimators are defined in Table 2.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Akemann, G., J. Baik, and P. Di Francesco (2011). *The Oxford handbook of random matrix theory*. Oxford University Press.

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & probability letters 80*(23), 1806–1813.

Amemiya, Y. and T. W. Anderson (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, 1453–1463.

Amenc, N., F. Goltz, L. Martellini, and P. Retkowsky (2011). Efficient indexation: An alternative to cap-weighted indices. *Journal Of Investment Management (JOIM), Fourth Quarter*.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Shi (2011). Estimating high dimensional covariance matrices and its applications. *Annals of Economics and Finance 12*(2), 199–215.

Barras, L. (2007). International conditional asset allocation under specification uncertainty. *Journal of Empirical Finance 14*(4), 443–464.

Barry, C. B. (1974). Portfolio analysis under uncertain means, variances, and covariances. *The Journal of Finance 29*(2), 515–522.

Behr, P., A. Guettler, and F. Truebenbach (2012). Using industry momentum to improve portfolio performance. *Journal of Banking & Finance 36*(5), 1414–1423.

Best, M. J. and R. R. Grauer (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *Review of Financial Studies 4*(2), 315–342.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance 54*(2), 655–671.

Brown, S. (1979). The effect of estimation risk on capital market equilibrium. *Journal of Financial and Quantitative Analysis 14*(2), 215–220.

Brown, S. J. (1989). The number of factors in security returns. *The Journal of Finance 44*(5), 1247–1262.

Candelon, B., C. Hurlin, and S. Tokpavi (2012). Sampling error and double shrinkage estimation of minimum variance portfolios. *Journal of Empirical Finance 19*(4), 511–527.

Chen, Y., A. Wiesel, and A. O. Hero (2011). Robust shrinkage estimation of high-dimensional covariance matrices. *Signal Processing, IEEE Transactions on 59*(9), 4097–4107.

Chopra, V. K. and W. T. Ziemba (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management 19*(2), 6–11.

Choueifaty, Y. and Y. Coignard (2008). Toward maximum diversification. *The Journal of Portfolio Management 35*(1), 40–51.

Clarke, R., H. d. Silva, and S. Thorley (2013). Risk parity, maximum diversification, and minimum variance: An analytic perspective. *The Journal of Portfolio Management 39*(3), 39–53.

Connor, G. and R. A. Korajczyk (1993). A test for the number of factors in an approximate factor model. *The Journal of Finance 48*(4), 1263–1291.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science 55*(5), 798–812.

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*(5), 1915–1953.

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

El Karoui, N. (2010). High-dimensionality effects in the markowitz problem and other quadratic programs with linear constraints: risk underestimation. *The Annals of Statistics 38*(6), 3487–3566.

Elton, E. J. and M. J. Gruber (1973). Estimating the dependence structure of share prices–implications for portfolio selection. *The Journal of Finance 28*(5), 1203–1232.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *the Journal of Finance 47*(2), 427–465.

Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147*(1), 186–197.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics 39*(6), 3320.

Fan, J., J. Zhang, and K. Yu (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association 107*(498), 592–606.

Frahm, G. and C. Memmel (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics 159*(2), 289–302.

Frost, P. A. and J. E. Savarino (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis 21*(3), 293–305.

Green, R. and B. Hollifield (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance 47*(5), 1785–1809.

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance 58*(4), 1651–1684.

Jobson, J. D. and B. Korkie (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association 75*(371), 544–554.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, 259–278.

Jorion, P. (1991). Bayesian and CAPM estimators of the means: Implications for portfolio selection. *Journal of Banking & Finance 15*(3), 717–727.

Kan, R. and G. Zhou (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis 42*(3), 621.

Klein, R. W. and V. S. Bawa (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics 3*(3), 215–231.

Kondor, I., S. Pafka, and G. Nagy (2007). Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking & Finance 31*(5), 1545–1573.

Kourtis, A., G. Dotsis, and R. N. Markellos (2012). Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. *Journal of Banking & Finance 36*(9), 2522–2531.

Laloux, L., P. Cizeau, M. Potters, and J.-P. Bouchaud (2000). Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance 3*(3), 391–397.

Ledoit, O. and M. Wolf (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance 10*(5), 603–621.

Ledoit, O. and M. Wolf (2004a). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management 30*(4), 110–119.

Ledoit, O. and M. Wolf (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*(2), 365–411.

Ledoit, O. and M. Wolf (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics 40*(2), 1024–1060.

Liu, Q. (2009). On portfolio optimization: How and when do we benefit from high-frequency data? *Journal of Applied Econometrics 24*(4), 560–582.

Maillard, S., T. Roncalli, and J. Teïletche (2010). The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management 36*(4), 60–70.

Marchenko, V. A. and L. A. Pastur (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik 114*(4), 507–536.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

Mazumder, R. and T. Hastie (2012). The graphical lasso: New insights and alternatives. *Electronic Journal of Statistics 6*, 2125–2149.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics 8*(4), 323–361.

Pantaleo, E., M. Tumminello, F. Lillo, and R. N. Mantegna (2011). When do improved covariance matrix estimators enhance portfolio optimization? an empirical comparative study of nine estimators. *Quantitative Finance 11*(7), 1067–1080.

Roll, R. (1977). A critique of the asset pricing theory's tests part i: On past and potential testability of the theory. *Journal of financial economics 4*(2), 129–176.

Roll, R. and S. A. Ross (1980). An empirical investigation of the arbitrage pricing theory. *The Journal of Finance 35*(5), 1073–1103.

Rosenow, B., V. Plerou, P. Gopikrishnan, and H. E. Stanley (2002). Portfolio optimization and the random magnet problem. *EPL (Europhysics Letters) 59*(4), 500–506.

Ross, S. A. (1976, December). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Sancetta, A. (2008). Sample covariance shrinkage for high dimensional dependent data. *Journal of Multivariate Analysis 99*(5), 949–967.

Santos, A. A., F. J. Nogales, E. Ruiz, and D. v. Dijk (2012). Optimal portfolios with minimum capital requirements. *Journal of Banking & Finance 36*(7), 1928–1942.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science 9*(2), 277–293.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance 19*(3), 425–442.

Stevens, G. V. (1998). On the inverse of the covariance matrix in portfolio analysis. *The Journal of Finance 53*(5), 1821–1827.

Trzcinka, C. (1986). On the number of factors in the arbitrage pricing model. *The Journal of Finance 41*(2), 347–368.

Tu, J. and G. Zhou (2011). Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics 99*(1), 204–215.

EDHEC Risk Institute, 393 Promenade des Anglais BP 3116, 06202 Nice Cedex 3 - France