Sparse Precision Matrices for Minimum Variance Portfolios

Gabriele Torri* Rosella Giacometti† Sandra Paterlini‡

June 6, 2018

Abstract

Financial crises are typically characterized by highly positively correlated asset returns due to the simultaneous distress on almost all securities, high volatilities and the presence of extreme returns. In the aftermath of the 2008 crisis, investors were prompted even further to look for portfolios that minimize risk and can better deal with estimation error in the inputs of the asset allocation models. The minimum variance portfolio à la Markowitz is considered the reference model for risk minimization in equity markets, due to its simplicity in the optimization as well as its need for just one input estimate: the inverse of the covariance estimate, or the so-called precision matrix. In this paper, we propose a data-driven portfolio framework based on two regularization methods, *glasso* and *tlasso*, that provide sparse estimates of the precision matrix by penalizing its L_1-norm. *Glasso* and *tlasso* rely on asset returns Gaussianity or t-Student assumptions, respectively. Simulation and real-world data results support the proposed methods compared to state-of-art approaches, such as random matrix and Ledoit-Wolf shrinkage.

Keywords: Minimum variance, precision matrix, graphical lasso, tlasso

JEL Codes: G11, C58

*gabriele.torri@unibg.it – University of Bergamo

†rosella.giacometti@unibg.it – University of Bergamo

‡sandra.paterlini@ebs.edu – EBS University, University of Trento

1 Introduction

Markowitz's mean-variance model (Markowitz, 1952) represents a cornerstone for asset allocation frameworks and financial theory in general (see for example Kolm et al. (2014) and references therein). Since then, many alternative methods and new strands of research have been established, fostered by the FinTech industry, which relies on the development of data-driven and automatic investment tools. Markowitz's simplistic Gaussian framework, the idea of risk-return optimization and diversification are still the starting points for the largest majority of more sophisticated recent approaches. Possibly, one of the main challenge of Markowitz's mean-variance model is the need to provide reliable estimates of the input parameters: the expected asset returns vector and the expected covariance matrix. Concerning the expected returns, it is acknowledged that they are extremely difficult to estimate and are often the main source for unreliable allocations (Michaud, 1989; Brodie et al., 2009), with suboptimal Sharpe ratios compared to the minimum variance portfolios (Black and Litterman, 1992). Most research has recently focused on building robust and reliable estimators for the covariance matrix to better control for the estimation error, especially to avoid its impact on asset weights estimates. In the minimum variance setting, the problem with the estimation of the covariance matrix is amplified by the fact that the input required for the analytical solution is its inverse, the precision matrix (Stevens, 1998). Hence, the optimal minimum variance portfolio strongly depends on the largest eigenvalues of the precision matrix, which are the smallest eigenvalues of the covariance matrix and are typically dominated by noise. As widely discussed in the literature, sample estimates often result in ill conditioned covariance matrices when the number of asset is relatively large compared to the depth of the time series (see e.g. Ledoit and Wolf, 2004b; Meucci, 2009; Won et al., 2013). This, in turn leads to optimal portfolios with extreme and unstable positions over time (see for instance Michaud, 1989 and Ledoit and Wolf, 2004a). Furthermore, the presence of positive multicollinearity among asset returns impacts even further the weight estimates by resulting in unrealistic short positions, which have to be offset by corresponding long positions. This is exacerbated when the number of assets is large compared to the number of observations available for the estimates. High volatilities, positive multicollinearity and the presence of extremes are typical during financial crises. Some works have focused on GARCH-based approaches to better capture time-varying volatility (see Engle, 2002). Here,

instead, we rely on a rolling-window mechanism to update the input estimates for the minimum variance portfolio. In the aftermath of the 2008 crisis, the academic literature has seen a surge of contributions to improve covariance estimation, such as the Ledoit-Wolf shrinkage estimator (Ledoit and Wolf, 2004a) and random matrix theory (Laloux et al., 1999). Other strands of research instead focus on robust optimization methods (DeMiguel and Nogales, 2009) and, most recently, on regularization methods such as lasso (Tibshirani, 1996), which relies on imposing a penalty function on the L_1-norm of the asset weight vector (DeMiguel et al., 2009a; Brodie et al., 2009; Fan et al., 2012). We refer the reader to Kremer et al. (2017) for a comparison of state-of-art techniques within a minimum risk framework.

Here, we introduce two approaches that rely on constraining the L_1-norm of the precision matrix to reduce the estimation error impact on optimal portfolio weights. The proposed methods are based on Markovian graphs to improve the stability of the precision matrix estimates. In particular, we consider two statistical set-ups: the first one assumes that asset returns are normally distributed, hereafter *glasso* (Friedman et al., 2008), while the second relies on the assumption of t-Student asset returns, which better fits assets returns that, as it is widely known, are leptokurtic and typically characterized by fat tails, hereafter *tlasso* (Finegold and Drton, 2011). To our knowledge, within the financial literature the *glasso* approach has been discussed in the context of portfolio optimization by Goto and Xu (2015) and Brownlees et al. (2015), while the *tlasso* has not received any attention yet. Here, we aim to fill some gaps in the literature by showing that *glasso* and *tlasso* are effective tools for the development of data-driven investment strategies. In particular, the main goal of the analysis is to investigate the out-of-sample performances of *glasso* and *tlasso* in an equity portfolio set-up. We focus on the minimum variance framework, since it has an analytical solution that depends on the precision matrix, allowing us to point out the effect of estimation error in the inputs. Moreover, aligned to many studies in the literature (e.g., DeMiguel et al., 2009a; Fan et al., 2012), we focus on the global minimum variance case, given that estimates of expected returns are typically unreliable (Michaud, 1989).

We underline that our focus here is on equity portfolios. Asset managers, whose mandate is investing on equity risky assets, might then benefit from our findings, when setting up investment strategies. Indeed, the identification of the global minimum variance portfolio is of interest also for investors who want to construct equity risky strategies by relying on quantitative modelling,

such as fintech companies. In particular, it has been shown that in presence of parameter uncertainty, the traditional two-funds portfolio, made of the risk-free asset and the tangency portfolio, is dominated by a three-fund portfolio that includes also the global minimum variance portfolio (Kan and Zhou, 2007).

The paper is structured as follow. Section 2 introduces the minimum variance portfolio framework and describes the *glasso* and *tlasso* approaches. Section 3 discusses the simulation set-up and the main results, while Section 4 provides empirical results on real-world financial data, when compared with state-of-art methods. Section 5 then concludes.

2 Methodology

2.1 The Minimum Variance Portfolio

Since the seminal work of Markowitz (Markowitz, 1952), the idea of risk minimization by diversification has become central to modern portfolio theory. Markowitz minimum variance portfolio framework is still considered the reference model for many scholars and practitioners. The simplicity of the risk diversification idea, quantified by linear dependence, resulting in the need to estimate "just" the covariance matrix, and the possibility to rely on an analytical solution, or to deal with a "simple" quadratic optimization problem, have been possibly among the main factors behind the widespread use of the minimum variance model, as stated below:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \\ s.t. \quad & \mathbf{1}'\mathbf{w} = 1, \end{aligned} \tag{1}$$

where $\boldsymbol{\Sigma}$ is the $n \times n$ true covariance matrix, $\mathbf{w}$ the $n \times 1$ vector of asset weights and $\mathbf{1}$ a $n \times 1$ unit vector. The optimization problem has then an analytical solution:

$$\mathbf{w}_{MV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}, \tag{2}$$

where $\mathbf{w}_{MV}$ is the vector of weights of the optimal minimum variance portfolio.

As $\boldsymbol{\Sigma}$ is unknown, an estimate $\widehat{\boldsymbol{\Sigma}}$ has to be computed to obtain the weights:

$$\hat{\mathbf{w}}_{MV} = \frac{\widehat{\mathbf{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}'\widehat{\mathbf{\Sigma}}^{-1}\mathbf{1}}. \tag{3}$$

Markowitz's minimum variance portfolios have been widely critized, mainly due to the sensitivity to estimation error in the input estimates, resulting in unrealistic extreme weights and bad out-of-sample properties. One of the main critiques is related to the use of the sample covariance estimates, and subsequently its inverse, as an input. Indeed such estimator has a slow convergence rate to the true covariance matrix. As a consequence, finite sample estimation is characterized by a higher eigenvalue dispersion compared to the true covariance matrix and by a low accuracy of the eigenvectors corresponding to the smallest eigenvalues, especially for matrices of large dimension estimated on short time series (Meucci, 2009). Another shortcoming is related to the role of the inverse of the covariance matrix $\mathbf{\Omega} \equiv \mathbf{\Sigma}^{-1}$, the so called *precision matrix*, in the solution of the optimization problem. From (2), it is clear that the accuracy of the estimation of the weights is directly related to the accuracy of the precision matrix. Using spectral decomposition, the relationship between the two matrices can be explicitly studied. In fact, the eigenvector decomposition of the covariance matrix can be expressed as $\mathbf{\Sigma} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$, where $\mathbf{V}$ is the matrix of eigevectors with $\mathbf{V}^{-1} = \mathbf{V}'$ and $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \cdots \geq \lambda_n$ is the diagonal matrix of the eigenvalues sorted in decreasing order on the main diagonal. Analogously, the eigendecomposition of the precision matrix is such that $\mathbf{\Omega} = \mathbf{U}\mathbf{\Delta}\mathbf{U}'$,where $\mathbf{\Delta} = diag(\delta_1, \ldots, \delta_n)$ with $\delta_1 \leq \cdots \leq \delta_n$ (note that in this case the eigenvalues are sorted in ascending order). By inverting the covariance matrix, we have

$$\begin{aligned} \mathbf{\Sigma}^{-1} &= (\mathbf{V}\mathbf{\Lambda}\mathbf{V}')^{-1} \\ &= (\mathbf{V}')^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1} \\ &= \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}'. \end{aligned} \tag{4}$$

It follows that $\mathbf{U} = \mathbf{V}$ and $\mathbf{\Delta} = \mathbf{\Lambda}^{-1}$ with the ith element $\delta_i = 1/\lambda_i$, that is, the eigenvectors of the precision matrix are the same as those of the covariance matrix and the eigenvalues are the reciprocal of those of the covariance matrix.

The consequence is that the dominant eigenvectors of the precision matrix (i.e., the ones corresponding to the largest eigenvalues) are the ones most likely influenced by noise and estimation error, especially in ill-conditioned covariance matrices.

2.2 Graphical Lasso or *glasso*

Graphical models can be useful to describe both the conditional and unconditional dependence structure of a set of variables. Gaussian Graphical Models (GGMs) are probably the most popular tools to capture the network structure of a set of variables. As Markowitz's model relies on the normality assumption of the asset returns, GGMs are the natural choice for capturing and estimating linear dependence among assets (see Dempster, 1972; Murphy, 2012). Such models are also known as *covariance selection* or *concentration graph* models, as they rely on the use of partial correlations as a measure of independence of any two variables, by exploiting the relationship between partial correlations and the inverse of the correlation matrix.

Let the asset return $X \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a random variable with a multivariate normal distribution with $\boldsymbol{\mu}$ the $n \times 1$ vector of the expected returns and $\boldsymbol{\Sigma}$ their $n \times n$ covariance matrix. We define the *precision matrix* as the inverse of the covariance matrix: $\boldsymbol{\Omega} \equiv \boldsymbol{\Sigma}^{-1}$.

We can then associate to the vector X an undirected graph defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes in the vertex set $\mathcal{V}$ correspond to each element of X, the edges $\mathcal{E}$ consist of the pairs of random variables with non-zero partial correlations: $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V} | r_{ij} \neq 0\}$, where r_{ij} denotes the partial correlation between two assets, that is, the correlation of the set of the remaining assets, but i and j. In the graph, the edge weights are the partial correlations. It can be shown that matrix of the partial correlations $\mathbf{R} = \{r_{ij}\}$ and the precision matrix $\boldsymbol{\Omega}$ are related as follows (Lauritzen, 1996):

$$\mathbf{R} = -\mathbf{D}\boldsymbol{\Omega}\mathbf{D}, \tag{5}$$

where $\mathbf{D} = diag(\frac{1}{\sqrt{\omega_{11}}}, \ldots, \frac{1}{\sqrt{\omega_{nn}}})$ and ω_{ii} is an entry of $\boldsymbol{\Omega}$.

The estimation of *sparse* precision matrices, that is, precision matrices with most off-diagonal entries exactly equal to zero, is then an important task. A common approach to introduce sparsity is to penalize the maximum

likelihood estimation problem by an L_1-norm. In the Gaussian case we can state the graphical lasso (*glasso*) estimation problem as

$$\widehat{\mathbf{\Omega}}_{glasso} = \arg\max_{\mathbf{\Omega}} \left(log|\mathbf{\Omega}| - tr(\mathbf{\Omega}\mathbf{S}) - \rho||\mathbf{\Omega}^{-}||_1 \right), \tag{6}$$

where $\mathbf{S}$ is the sample covariance matrix, ρ is a tuning parameter that controls the sparsity (i.e. the larger ρ, the larger the number of elements of the precision matrix set equal to zero), $|\cdot|$ is the determinant, $tr(\cdot)$ the trace, $||\cdot||_1$ the L_1-norm of a matrix and $\mathbf{\Omega}^{-}$ is a square $n \times n$ matrix with the off-diagonal elements equal to the corresponding elements of $\mathbf{\Omega}$ and the diagonal elements equal to zero (Friedman et al., 2008).[1]

As pointed out by Yuan and Lin (2007), the use of an L_1 penalty allows to induce sparsity in the precision matrix, making it possible to perform model selection and parameter estimation simultaneously. The choice of the tuning parameter ρ, that controls the level of sparsity of the estimates assumes therefore a major role in the estimation and it will be examined in Section 3.3.

The statistical properties of the *glasso* estimator have been studied, among others, by Banerjee et al. (2008), Rothman et al. (2008) and Lam and Fan (2009). Rothman et al. (2008) show that under some regularity conditions and for a suitable choice of the parameter ρ, the estimator in (6) has a rate of convergence to the true parameter $\mathbf{\Omega}$ in the Frobenius norm of order $\sqrt{((n + k_{\mathbf{\Omega}}) \log(n)/t)}$, where $k_{\mathbf{\Omega}}$ is the number of non-zero off diagonal entries of the true matrix $\mathbf{\Omega}$ and t the number of observations. The main implication is that the convergence is faster for matrices that are truly sparse. Lam and Fan (2009) studied *glasso*'s consistency and *sparsistency* (i.e., the property that all parameters that are zero are actually estimated as zero with probability tending to one, also known in the literature as *selection consistency*). They show that sparsistency requires a lower bound on the rate of the regularization parameter ρ, while an upper bound is necessary to control the bias introduced by the L_1 penalty and to obtain a consistent estimator.[2] Under some technical conditions, the compatibility of these bounds requires the number of off-diagonal non-zero entries of the true precision matrix to

[1] The original specification proposed by Friedman et al. (2008) applied the penalty to the entire matrix $\mathbf{\Omega}$. The version of the model with the penalty applied to $\mathbf{\Omega}$ is the one studied by Rothman et al. (2008) and is currently implemented in the R package *'glasso'* (Friedman et al., 2014).

[2] See Theorem 2 and Technical Condition (B) in Lam and Fan (2009).

be no larger than $O(n)$ (Lam and Fan, 2009). In practical terms, the true precision matrix has to be sparse enough, and the *glasso* estimates of dense precision matrices will not be consistent. On the other hand, the modelization of sparse precision matrices has the advantage of reducing the variability of the estimates, as it will be discussed in Section 2.3 with reference to the *regression hedge*.

The optimization problem (6) is convex, as proved by Banerjee et al. (2008). Friedman et al. (2008) proposed an efficient algorithm to solve it, that is briefly described in Appendix A.

Empirical analysis show that *glasso* estimates are better conditioned compared to the sample covariance matrix, even when the number of covariates n is close to the number of observations t (see e.g., Goto and Xu, 2015). Moreover, the solution $\widehat{\mathbf{\Omega}}_{glasso}$ is always unique and has bounded eigenvalues,[3] also when $n \leq t$, allowing the use this method also in high-dimensional setting, in which the sample covariance matrix estimate is singular. Finally, we point out that the sparsity of the precision matrix does not necessariliy correspond to the sparsity of the covariance matrix.

2.3 Sparse Precision Matrix Estimation and Regression Hedge

From a financial point of view, the sparsity of the precision matrix can be considered in the framework of *regression hedge*. In fact, as discussed by Stevens (1998) and Goto and Xu (2015), the precision matrix has an interpretation in terms of optimal hedging between assets: specifically, the ith row (or column) of $\mathbf{\Omega}$ is proportional to the ith asset's hedge portfolio. Such ith hedge portfolio consists of the combination of a long position in the ith asset and $(n-1)$ positions in all the other assets that minimize the variance of the tracking error of the ith asset w.r.t. the remaining $(n-1)$ assets. The ith tracking portfolio is defined as follows:

$$X_{i,\tau} = \alpha_i + \sum_{k=1,k\neq i}^{n} \beta_k^{(i)} X_{k,\tau} + \varepsilon_{i,\tau} \qquad i = 1, \ldots, n, \tag{7}$$

where $X_{i,\tau}$ is the ith asset return at time τ, $\beta_k^{(i)}$ is the coefficient of asset k in the regression for asset i, $\varepsilon_{i,\tau}$ is the unhedgeable component of $X_{i,\tau}$.

[3] see Banerjee et al. (2008), Theorem 1.

The regressions in (7) are typically defined in the financial literature as *regression hedge*. As shown in Stevens (1998), the OLS estimates of the βs can be easily related to the precision matrix. We identify the following partition of the sample covariance matrix $\mathbf{S}$

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{\setminus i,\setminus i} & \mathbf{s}_{\setminus i,i} \\ \mathbf{s}'_{\setminus i,i} & s_{i,i} \end{pmatrix}, \tag{8}$$

where $X_{\setminus i}$ denotes all the elements of X except the ith, $\mathbf{S}_{\setminus i,\setminus i}$ denotes the firsts $n-1$ rows and columns of $\mathbf{S}$, $\mathbf{s}_{\setminus i,i}$ the first $n-1$ elements of the last column, and $s_{i,i}$ is the element in the last row and column.[4]

We then have that the following relationship holds:

$$\hat{\beta}^{(i)} = \mathbf{S}^{-1}_{\setminus i,\setminus i}\mathbf{s}_{\setminus i,i}, \tag{9}$$

where $\hat{\beta}^{(i)}$ is the $((n-1)\times 1)$ vector of the coefficients in the ith regression hedge. Moreover, let $v_i = \text{var}(\varepsilon_i)$ be the variance of the residual ε_i, then the elements of $\widehat{\mathbf{\Omega}}$ can then be computed as follows:[5]

$$\hat{\omega}_{ij} = \begin{cases} -\dfrac{\hat{\beta}^{(i)}_j}{v_i} & \text{if } i \neq j \\ \dfrac{1}{v_i} & \text{if } i = j \end{cases}. \tag{10}$$

Further details can be found in Stevens (1998).

In financial applications, the regression hedge framework generally suffers in presence of multicollinearity among the regressors. Regularization techniques that allow to set some βs to zero, can then provide more reliable estimates and better out-of-sample performances, at the cost of introducing some bias. One of the most common techniques is the lasso regression, that introduces an L_1-norm penalty in the estimation problem. As we discuss now, the *glasso* estimator allows to introduce an L_1 regulatization of all the regression hedges, while maintaining the relationship with the precision matrix as presented in (10).

The naive application of the lasso penalty on each regression hedge, indeed is not consistent with (10), since this approach does not constrain $\widehat{\mathbf{\Omega}}$ to

[4]Notice that this representation implies a permutation of the rows and columns to have the ith asset as the last one.

[5]v_i can be interpreted as the unhedgeable component of $X_{i,t}$.

be symmetric and positive definite. Instead, the glasso algorithm estimates all the regression hedges iteratively as n coupled lasso problems. The information is shared between the lasso problems through the common estimate of the matrix $\mathbf{G}$, providing a positive definite, symmetric and sparse estimate of $\mathbf{\Omega}$ (Friedman et al., 2008). This allows to extend the analysis of Stevens (1998) to the sparse case (Goto and Xu, 2015). On one hand, the use of the lasso penalization in the regression hedge equations introduces a bias. On the other hand it reduces the estimation variation, leading to a more efficient estimator of the precision matrix.

Overall, the *glasso* method has a shrinkage effect on the βs of the regression hedge, filtering the estimation noise in $\mathbf{\Sigma}$ and its effect when computing $\mathbf{\Omega} \equiv \mathbf{\Sigma}^{-1}$. Empirical evidence suggests that the *glasso* estimates of $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ are better conditioned than the sample covariance matrix (Goto and Xu, 2015). The spectrum of *glasso* estimates is therefore typically less disperse than the one of sample covariance.

Despite its appealing properties, so far, we are aware of only two applications of *glasso* within asset allocation frameworks (Goto and Xu, 2015; Brownlees et al., 2015). Here, we contribute to the literature by providing further evidence when comparing *glasso* to state-of-art methods. Moreover, as widely known, asset returns normality assumption is too stringent, as stylized facts suggest that asset returns have a leptokurtic distribution, which can be better captured by a t-Student assumption (Cont, 2001). Hence, we move one step further by introducing the so-called *tlasso* model that allows to estimate the precision matrix under the assumption of multivariate t-Student distribution of asset returns.

2.4 Robust Graphical Modeling With *tlasso*

As widely discussed in the statistical literature, deviations of returns from Gaussianity can significantly impact the estimation and the inference on GGMs. Asset return distributions typically deviate from normality by having fatter tails and leptokurtic distributions. Hence, the t-Student assumption with a low number of degrees of freedom is considered a better choice to model asset returns. Moreover, relying on such distribution can provide more robust estimates in presence of outliers or contaminated data (Lange et al., 1989). Recently, Finegold and Drton (2011) introduced the so called *tlasso*, replacing the *glasso* Gaussian assumption with a t-Student to provide a tool for robust model selection. The *tlasso* algorithm estimates then a sparse

precision matrix under the assumption that the data follow a multivariate t-Student distribution.

Let $X = (X_1, \ldots, X_n)$ be a random vector following a multivariate t-Student distribution $t_n(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, df)$, with df degrees of freedom, mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Psi}^{-1}$ ($n \times n$ positive semi-definite matrix). The covariance matrix is then

$$\boldsymbol{\Sigma} = \frac{df}{df - 2}\boldsymbol{\Psi}^{-1}, \tag{11}$$

and the precision matrix is

$$\boldsymbol{\Omega} = (\boldsymbol{\Sigma})^{-1} = \frac{df - 2}{df}\boldsymbol{\Psi}. \tag{12}$$

Similarly to the Gaussian case, we can associate a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ in which $\mathcal{E} = \{(i, j) \in \mathcal{V} \times \mathcal{V} | r_{ij} \neq 0\}$ and the edge weights are the corresponding partial correlations r_{ij} computed from the precision matrix.

Under the t-Student assumption, in contrast to the Gaussian set-up, the absence of correlation does not necessarily correspond to conditional independence (Baba et al., 2004). However, despite the lack of conditional independence for $\omega_{ij} = 0$ (where ω_{ij} is an element of $\boldsymbol{\Omega}$), we have that, if two nodes j and k are separated by a set of nodes C in $\mathcal{G}$, then X_j and X_k are conditionally uncorrelated given $X_{\{C\}}$ (see Finegold and Drton, 2011, Proposition 1). Disconnected vertices can be considered orthogonal to each other after the effects of other variables are removed. The absence of conditional correlations entails that a mean-square error optimal prediction of variable X_j can be based on the variables X_k, which correspond to neighbours of the node j in the graph.

We adopt the estimation procedure introduced by Finegold and Drton (2011), that exploits the scale-mixture representation of the multivariate t-Student distribution consisting of a multivariate Gaussian and a gamma distribution (Kotz and Nadarajah, 2004) and uses an EM-algorithm (Expectation-Maximization). In particular, the E-step consists in the estimation of the mixing gamma variable and the M-step in the estimation of parameters $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Psi}}$ given the latent variable (the degrees of freedom df are assumed to be known in this version of the algorithm). Since the sparse parameter $\boldsymbol{\Psi}$ is the precision matrix of the conditional Gaussian variable, it can be estimated efficiently in the M-step of the algorithm using *glasso* (Finegold and Drton, 2011). The E- and the M-steps are then iterated until convergence. The

estimate of the precision matrix $\widehat{\mathbf{\Omega}}_{tlasso}$ of the multivariate t-Student vector is finally obtained by rescaling the estimate $\widehat{\mathbf{\Psi}}_{tlasso}$ using (12).

The *tlasso* procedure is computationally efficient since it is based on *glasso* algorithm at every M-step. While convergence to a stationary point is guaranteed in the penalized versions of EM (McLachlan and Krishnan, 1997), the algorithm is not guaranteed to converge to the global maximum since the *tlasso* penalized log-likelihood function to be maximized is not concave (Finegold and Drton, 2011).

The scale-mixture representation of the multivariate t-Student also allows the regression hedge interpretation as in the Gaussian case. Indeed, the non penalized version of the EM algorithm can be interpreted as an iteratively reweighted least square estimation of the regression of each variable on all the others, as shown in Lange et al. (1989). In the penalized case, the estimation is consistent with the iteratively reweighted *glasso* estimation.

Finally, we underline that in the empirical application we use the *tlasso* and *glasso* to estimate the correlation matrix and its inverse (rather than the covariance and precision matrices). We then obtain the corresponding estimates of the covariance and precision matrix by scaling the output using the sample variances. This approach ensures that the penalization is not influenced by the scale of the variables (Højsgaard et al., 2012) and, as proved by Rothman et al. (2008), ensures a faster convegence of the estimator in the matrix 2-norm.

3 Simulation Analysis

We conduct a simulation analysis to test the empirical properties of the precision matrix estimates by *glasso* and *tlasso*. In particular, our two main goals are to measure the quality of the estimates of the true covariance and precision matrices and to assess their impact on the solution of the minimum variance portfolio. We compare the results obtained with *glasso* and *tlasso* to the traditional sample covariance matrix, as well as to the naive equally weighted portfolio strategy (EW) and to two state-of-art covariance estimation methods: random matrix theory filtering (RMT) (Bouchaud and Potters, 2009) and Ledoit Wolf shrinkage estimation (LW) (Ledoit and Wolf, 2004a). Appendix B provides a brief description of these methods with relevant references.

3.1 Statistical Performance Measures

To test the quality of the covariance and precision matrix estimates we measure the error, bias and inefficiency with respect to the true parameters. For explanatory purposes, we describe the measures referring to the covariance matrix $\mathbf{\Sigma}$, which can then be computed also for the precision matrix $\mathbf{\Omega}$.

First, we introduce the following loss function:

$$\text{Loss}[\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}] \equiv ||\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}||^2, \tag{13}$$

where $|| \cdot ||^2 = tr\big[(\cdot)^2\big]$ is the square of the Frobenius norm. Then, we can compute three measures to quantify the estimation accuracy. First, the *error*, that is the square root of the expected loss between the estimated and the true parameters:

$$\text{Err}[\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}] = \sqrt{\mathbb{E}\Big[||\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}||^2\Big]};$$

Second, the *inefficiency*, which is a measure of dispersion of the estimates and is computed as:

$$\text{Inef}[\widehat{\mathbf{\Sigma}}] = \sqrt{\mathbb{E}\Big[||\mathbb{E}\big[\widehat{\mathbf{\Sigma}}\big] - \widehat{\mathbf{\Sigma}}||^2\Big]};$$

Finally, the *bias*, that quantifies the distance between the expected value of the estimated covariance and the true parameter:

$$\text{Bias}[\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}] = \sqrt{||\mathbb{E}\big[\widehat{\mathbf{\Sigma}}\big] - \mathbf{\Sigma}||^2}.$$

As widely known, the following relationship holds:

$$\text{Err}^2[\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}] = \text{Bias}^2[\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}] + \text{Inef}^2[\widehat{\mathbf{\Sigma}}]. \tag{14}$$

Using regularization and shrinkage techniques, we expect to reduce the estimation error by increasing the efficiency of the estimator compared to the sample covariance, typically at the cost of an increased bias.

Glasso and *tlasso* rely on the direct regularization of the precision matrix. Therefore, we expect them to provide good estimates of the optimal assets' weights in the minimum variance portfolio framework, given that the precision matrix represents the input of the optimization. To evaluate the overall

impact of the estimation error in $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Omega}}$, we compute the empirical, actual and oracle risk of optimal portfolios. In particular, considering the standard deviation as a risk measure we have:

$$R_{\text{empirical}} = \sqrt{\hat{\mathbf{w}}_{MV}^T \widehat{\mathbf{\Sigma}} \hat{\mathbf{w}}_{MV}}, \tag{15}$$

$$R_{\text{actual}} = \sqrt{\hat{\mathbf{w}}_{MV}^T \mathbf{\Sigma} \hat{\mathbf{w}}_{MV}}, \tag{16}$$

$$R_{\text{oracle}} = \sqrt{\mathbf{w}_{MV}^T \mathbf{\Sigma} \mathbf{w}_{MV}}, \tag{17}$$

where $\mathbf{\Sigma}$ is the true covariance matrix, $\widehat{\mathbf{\Sigma}}$ is an estimate, $\mathbf{w}_{MV}$ is the optimal vector of minimum variance weights with $\mathbf{\Sigma}$ as input and $\hat{\mathbf{w}}_{MV}$ is the optimal weight vector for $\widehat{\mathbf{\Sigma}}$.

These measures give us insights on the impact of the estimation error of $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Omega}}$ in the optimization process: the empirical risk represents the perceived risk by the investor, the actual risk is the one which the investor is exposed to, while the oracle risk is the minimum risk possible given the true covariance matrix (Fan et al., 2012). Since in the real world the last two are unknown, the estimation process should minimize errors due to estimation and provide an empirical risk as close as possible to the oracle and actual.

3.2 Simulation Set-up

We consider two different approaches for the simulation set-up. The first one is a three-factors model, similar to the one in Fan et al. (2012) (which we denote as the *Factor Model* data). It assumes that the excess returns of the assets are generated according to:

$$X_i = b_{i1} f_1 + b_{i2} f_2 + b_{i3} f_3 + \varepsilon_i \qquad i = 1, \ldots, n, \tag{18}$$

where f_1, f_2 and f_3 are the three factors' returns, b_{ik} are the factor loadings for the kth factor and for the ith asset and ε_i is the idiosyncratic noise. The factors' are generated by a multivariate t-Student distribution with 5 degrees of freedom,[6] while the idiosyncratic terms are generated from a univariate t-Student distribution. The parameters of factor returns, factor loadings and level of idiosyncratic noise are calibrated on real market data (see Fan et al., 2012 for more details).

[6]In the original model the factors followed a multivariate normal distribution (Fan et al., 2012). We used a t-Student to capture the leptokurtic distribution of financial time series (Cont, 2001).

In the second approach (henceforth *Simulated S&P100*), we generate the assets' returns by a multivariate t-Student distribution with 5 degrees of freedom and a covariance matrix estimated on the daily returns of constituents of the S&P100 equity index for the period 01/01/2006 – 31/12/2016, adding to each asset a noise factor distributed as univariate t-Student with variance equal to 0.1 times the variance of each asset.

For each setting, we consider two configurations characterized by different dimensionality: 50 and 85 assets, respectively. In both cases the parameters are estimated over a window of 100 observations and we consider 30 simulation runs. The number of degrees of freedom for the *tlasso* has been set equal to 5.

3.3 Optimal Choice of ρ

The structure of the precision matrix estimated by *glasso* and *tlasso* depends largely on the choice of the penalization parameter ρ, that controls the level of sparsity in the precision matrix. We select the optimal ρ on a grid of values using the Bayesian Information Criterion (BIC)

$$BIC = -2log(Lik_{\rho_i}) + k_{\widehat{\mathbf{\Omega}}} \times log(t), \tag{19}$$

where Lik_{ρ_i} is the value of the likelihood function corresponding to the ith value of ρ in the grid, $k_{\widehat{\mathbf{\Omega}}}$ is the number of non-null elements in the estimate of the precision matrix, and t the number of observations.[7] The grid is composed by 20 logarithmically spaced values between 0 and 1. The choice of this interval guarantees that the *glasso* estimates span from a completely dense precision matrix ($\rho = 0$) to a completely sparse one ($\rho = 1$) when estimated using the correlation matrix as input.[8]. For *tlasso* such result is not guaranteed, but we found empirically that such interval is wide enough to include the estimate characterized by the optimal BIC in all the cases (see Figures 1 and 2).

For sake of brevity, we report exclusively the parameter calibration of *tlasso*; the procedure and the results are analogous for *glasso* and available

[7] In the case of *glasso* we refer to the likelihood of a multivariate normal distribution, while with *tlasso* we refer to the one of a multivariate t-Student distribution.

[8] The result follows from Corollary 1 in Witten et al. (2011), according to which the ith node is fully unconnected to all other nodes if and only if $|\mathbf{\Sigma}_{ij}| \leq \rho \quad \forall i \neq j$. When $\mathbf{\Sigma}$ is the correlation matrix, all its elements are smaller or equal to one and therefore for $\rho = 1$ all the elements are disconnected, that is, the precision matrix is diagonal.

from the authors upon request.

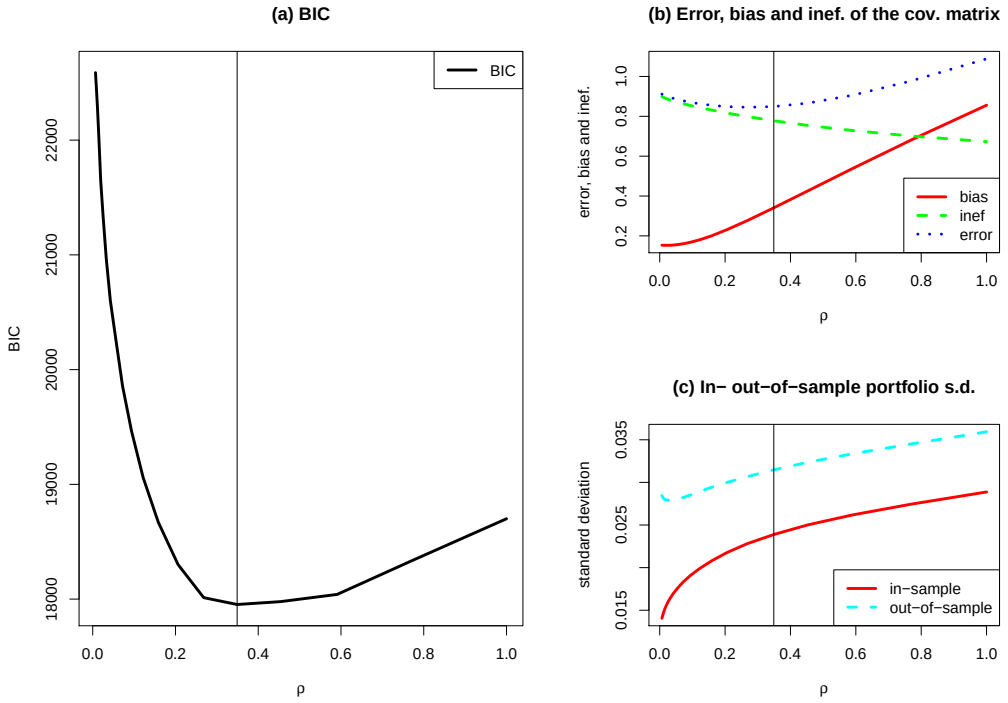

Figure 1: Optimal Choice of ρ for *tlasso*, *Factor Model*, 85 assets. Panel (a) displays the value of BIC for every value of ρ in the grid. Panel (b) shows the values of error, bias and inefficiency of the estimation of the covariance matrix (see (13) and (14)). Panel (c) displays the in- and out-of-sample standard deviation of the optimal minimum variance portfolios. The value reported are based on 30 runs and the vertical lines denote the median of the optimal ρ across the runs.

Figures 1 and 2 show for *tlasso* the values of error, bias and inefficiency of the covariance matrix estimator, as well as the in- and out-of-sample standard deviation as a function of ρ for the *Factor Model* and *Simulated S&P100*, respectively. In both cases we consider 85 assets estimated on a window of 100 observations. Panel (a) reports the value of BIC, Panel (b) the error, bias and inefficiency of the estimation of covariance matrix and Panel (c) the in-sample and out-of-sample standard deviations. We observe in Panel (a) that the BIC

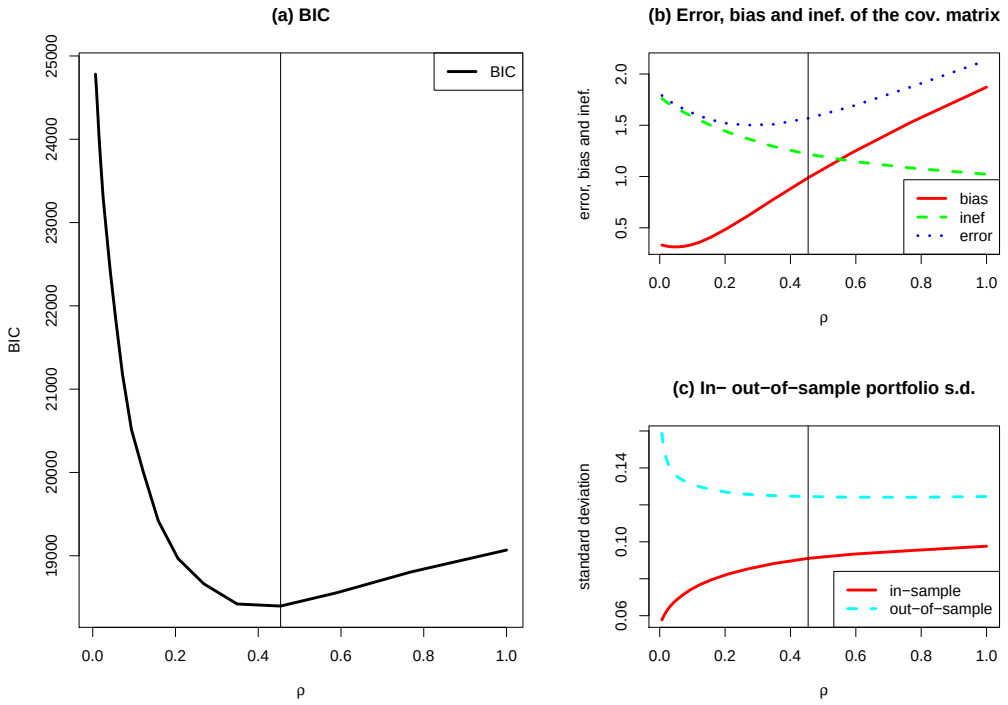

Figure 2: Optimal Choice of ρ for *tlasso*, *Simulated S&P100*, 85 assets. Panel (a) displays the value of BIC for every value of ρ in the grid. Panel (b) shows the values of error, bias and inefficiency of the estimation of the covariance matrix (see (13) and (14)). Panel (c) displays the in- and out-of-sample standard deviation of the optimal minimum variance portfolios. The value reported are based on 30 runs and the vertical lines denote the median of the optimal ρ across the runs.

optimal model lies within the interval $\rho \in [0, 1]$, in both cases with values close to 0.4, characterized by a medium sparsity level in $\widehat{\mathbf{\Omega}}$. From Panel (b), the bias increases with the values of ρ, while the inefficiency decreases. This pattern is consistent with the fact that, for the *tlasso* estimates computed with higher values of ρ, the number of parameters to estimate is smaller, given that more elements of the precision matrix are set equal to zero. The overall estimation error reaches a minimum for intermediate values of ρ not distant from the ones chosen by the BIC. Concerning the portfolio performances,

we see from Panel (c) that the effect of ρ on the out-of-sample standard deviation is different for the two simulation settings: in the case of *Factor Model*, the portfolios have minimum risk with a rather low value of ρ, while for *Simulated S&P100*, the out-of-sample risk is minimized for a wider range of ρ.

3.4 Simulation Results

3.4.1 Accuracy of the Estimates

Table 1 reports bias, inefficiency and error for the covariance (columns 2–4) and the precision matrix (columns 5–7) computed on 30 runs for four test cases: *Factor Model* and *Simulated S&P100*, with 50 and 85 assets. It also reports the average condition numbers of the estimates for 30 runs (column 8) and the ones of the true covariance matrices. Concerning the estimation of the covariance matrix, we observe that *glasso*, *tlasso*, RMT and LW present in all the cases a low inefficiency compared to the sample covariance matrix. This comes at the cost of a higher bias. As a consequence of these two opposite effects, the overall error levels end up being similar. Indeed, *glasso*, *tlasso* and LW provide only minor improvements in terms of overall error with respect to the naive sample covariance approach, while RMT shows in three out of four test cases an estimation error larger than the sample covariance due to a particularly high bias. The results are consistent for all the simulation set-ups. The fact that the *Factor Model* test cases are characterized by larger errors then the *Simulated S&P100* can be explained by higher values of the entries of the covariance matrix, resulting from higher volatility and collinearity in the data.

In the context of minimum variance portfolio selection, the focus is on the estimation of the precision matrix, that is the input of the closed form optimal solutions in (3). Indeed, the analysis of the estimation error of such matrix displays a rather different picture, more aligned to the well documented pitfalls of minimum variance portfolios estimated using sample covariance (Michaud, 1989). The estimation of the precision matrix obtained by inverting the sample covariance is indeed characterized by an error much higher than the alternative estimates in all the test cases considered, especially when the number of assets is large. For instance, in Panel 2 (*Factor Model* with $n = 85$), the error for the precision matrix is equal to 102120.68, while the error of the estimates obtained using *glasso* and *tlasso* is equal to

less than a tenth of it: 8928.08 and 8985.56, respectively. The large error of the sample covariance estimator is not surprising, given that such matrices are characterized by high condition numbers, and their inverses are therefore highly sensitive to estimation error (the most relevant eigenvectors of the precision matrix are indeed the ones corresponding to the smallest eigenvalues of the covariance matrix, which are typically dominated by noise). *Glasso* and *tlasso* show good performances, with the smallest error in test case 1, 3 and 4. The difference with RMT and LW in terms of error are generally moderate. The results are similar for the *Factor Model* and the *Simulated S&P100* set-ups, although in the latter the absolute value of the estimation errors for the precision matrix are smaller than in the *Factor Model*. This difference is probably due to the worse conditioning of the covariance matrices in the *Factor Model* cases, which amplifies the estimation error of the covariance matrix.

3.4.2 Empirical, Actual and Oracle Risk

Table 2 reports empirical, actual and oracle risk. Concerning the *Factor Model* test cases (Panel 1 and 2), we see that, in terms of actual risk, *glasso* and *tlasso* obtain the best results: in the case with 50 assets, they have an actual risk of 0.038 and 0.037 (Panel 1), lower than the LW, RMT and EW portfolios. They are also the portfolios with the lowest actual risk in the setting with 85 assets (Panel 2). The portfolios estimated using the sample covariance are a peculiar case, since they obtain the lowest actual risk for 50 assets, but have the highest actual risk for 85 assets. This is consistent with the high error in the estimation of the precision matrix, as highlighted in Section 3.4.1. Such portfolios are also the ones with the largest difference between actual and empirical risk (especially in Panel 2, 85 assets settings, where they are equal to 0.045 and 0.006, respectively), while the other techniques provide much less divergent values (in Panel 2 for instance the actual and empirical risk for *tlasso* are equal to 0.030 and 0.032). As the empirical risk is the only one known to investors in real-world applications, positive differences between actual and empirical can lead to risk underestimation.

In the *Simulated S&P100* test cases we observe that, again, *glasso* and *tlasso* portfolios have the lowest actual risk. The advantage over RMT and LW in this case is limited, much smaller than the *Factor Model* framework (Panels 3 and 4). Sample covariance portfolios show a particularly high actual risk (0.182 and 0.270 for the settings with 50 and 85 assets, respectively,

Table 1: Bias, inefficiency and error for the estimation of covariance and precision matrices; condition number of covariance matrix. Values are computed in 30 runs.

Panel 1 - *Factor Model* - 50 assets, 100 obs. (cond. number of $\mathbf{\Sigma}$: 12359.42)							
	Covariance matrix			Precision matrix			Cond. number
	bias	inef	error	bias	inef	error	
sample cov.	0.08	0.64	0.64	8390.62	4801.40	9667.26	29770.64
glasso	0.19	0.60	0.63	5476.29	190.52	5479.60	1401.60
tlasso	0.16	0.57	0.59	5331.74	244.77	5337.35	1645.16
RMT	0.73	0.43	0.85	6127.47	79.91	6127.99	425.54
LW	0.21	0.60	0.63	5579.33	327.75	5588.94	1336.68
Panel 2 - *Factor Model* - 85 assets, 100 obs. (cond. number of $\mathbf{\Sigma}$: 27016.97)							
	Covariance matrix			Precision matrix			Cond. number
	bias	inef	error	bias	inef	error	
sample cov.	0.12	0.96	0.97	74698.65	69632.92	102120.68	296622.82
glasso	0.32	0.89	0.94	8924.83	240.91	8928.08	2119.29
tlasso	0.35	0.82	0.89	8981.38	274.12	8985.56	1938.62
RMT	1.27	0.59	1.41	9706.34	100.23	9706.86	679.27
LW	0.30	0.90	0.95	8732.19	702.21	8760.37	2825.45
Panel 3 - *Simulated S&P100* - 50 assets, 100 obs. (cond. number of $\mathbf{\Sigma}$: 1064.35)							
	Covariance matrix			Precision matrix			Cond. number
	bias	inef	error	bias	inef	error	
sample cov.	0.20	1.20	1.22	402.24	601.64	723.72	2014.03
glasso	0.47	1.07	1.17	263.69	57.79	269.94	275.73
tlasso	0.87	0.75	1.15	284.75	45.29	288.32	190.13
RMT	1.25	0.73	1.45	312.06	31.60	313.66	137.84
LW	0.53	0.92	1.06	289.43	64.61	296.56	294.64
Panel 4 - *Simulated S&P100* - 85 assets, 100 obs. (cond. number of $\mathbf{\Sigma}$: 1630.18)							
	Covariance matrix			Precision matrix			Cond. number
	bias	inef	error	bias	inef	error	
sample cov.	0.77	3.78	3.85	4327.64	9170.36	10140.22	37909.67
glasso	1.21	3.42	3.63	309.70	79.65	319.77	446.39
tlasso	1.49	1.58	2.17	330.86	71.30	338.46	315.02
RMT	2.13	2.35	3.17	367.85	50.81	371.34	218.32
LW	0.99	1.82	2.07	332.10	103.70	347.91	428.29

against the 0.135 and 0.122 for *tlasso*), highlighting the limits of this simple estimation technique. Equally weighted portfolios, typically considered a difficult benchmark to beat (DeMiguel et al., 2009b), do not seem to show interesting performances in terms of the risk measures that we consider, exhibiting the highest actual risk in three out of four test cases.

Summing up, *glasso* and *tlasso* perform well in all the test cases, exhibiting low portfolio risk exposures compared to the alternative techniques, especially in markets characterized by ill-conditioned covariance matrices (i.e. the *Factor Model* case). This suggests that these techniques might be particularly suitable in presence of multicollinearity, a characteristic typically associated to financial crises.

Table 2: Empirical, actual and oracle risk for the optimal portfolios. Average results over 30 runs.

Panel 1 - *Factor Model*						
	50 assets, 100 observations			85 assets, 100 observations		
	empirical	actual	oracle	empirical	actual	oracle
sample cov.	0.015	0.032	0.022	0.006	0.045	0.018
glasso	0.039	0.038	0.022	0.031	0.030	0.018
tlasso	0.038	0.037	0.022	0.032	0.030	0.018
RMT	0.052	0.053	0.022	0.042	0.041	0.018
LW	0.039	0.044	0.022	0.028	0.037	0.018
EW	0.190	0.193	0.022	0.186	0.188	0.018
Panel 2 - *Simulated S&P100*						
	50 assets, 100 observations			85 assets, 100 observations		
	empirical	actual	oracle	empirical	actual	oracle
sample cov.	0.074	0.182	0.122	0.032	0.270	0.103
glasso	0.098	0.136	0.122	0.079	0.123	0.103
tlasso	0.101	0.135	0.122	0.082	0.122	0.103
RMT	0.100	0.136	0.122	0.080	0.125	0.103
LW	0.102	0.138	0.122	0.080	0.129	0.103
EW	0.239	0.244	0.122	0.217	0.228	0.103

4 Real-World Data Analysis

4.1 Empirical Set-up

We test the performance of the portfolio determined by *glasso* and *tlasso* on four real-world datasets. The first two are provided by Kenneth French and are publicly available on his website.[9] One consists of the monthly returns of 48 US industry portfolios (FF 48) and the other of the returns of 100 portfolios formed on size and book-to-market ratio (FF 100) of US companies. The third and the fourth datasets are made of the stock returns of the constituents of S&P 100, at monthly and daily frequency (S&P 100 monthly and S&P 100 daily). The choice of the datasets spans different combinations of constituents, dimensionality and sampling frequency, providing robustness to the results.

We analyse the out-of-sample performances using a rolling window approach, rebalancing the portfolios every three months by computing the optimal global minimum variance portfolio on a window of fixed size. The out-of-sample period is defined for all the portfolios from January 2006 to December 2016. The estimation windows consist of 1 year (252 observations) in the case of the daily data, while for the monthly data we considered longer time windows in order to have sufficient data points (10 years, 120 observations). The main characteristics of the datasets are summarized in Table 3.

We evaluate the resulting portfolios in terms of risk/return profile (computing standard deviation, average return and Sharpe ratio) and in terms of portfolio composition, computing statistics relative to shorting, diversification and turnover. As in the simulation study, we estimated the *tlasso* with 5 degrees of freedom.

4.2 Empirical Results

Table 4 displays the performance measures for the portfolios estimated on real data. As we construct minimum variance portfolios, we focus in particular on the standard deviation, which is the quantity of interest in the optimization. Still, we also compute the average return and the Sharpe ratio of the portfolios to analyse the risk-adjusted return profiles.

[9] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Table 3: Descriptive statistics of the real-world datasets. The first three columns report the number of assets (n), the window size for the calibration (t) and the ratio between these two values (n/t), respectively. Columns 4 and 5 report the period spanned by each dataset and the frequency of the data. Note that, concerning the S&P100, we included in the analysis only the assets whose time series spanned the entire time period.

Portfolio	n	t	n/t	time period	data freq.
FF 48	48	120	0.40	01/1996 - 12/2016	monthly
FF 100	100	120	0.83	01/1996 - 12/2016	monthly
S&P 100 monthly	86	120	0.72	01/1996 - 12/2016	monthly
S&P 100 daily	91	252	0.36	01/2005 - 12/2016	daily

As expected, the out-of-sample standard deviation is larger than the in-sample in all the cases. The difference between the two is especially relevant for the portfolios computed using the sample covariance estimator when the number of assets is large compared to the length of the estimation window. For instance, in the FF 100 case, in which we consider 100 assets and the estimation window is made of 120 observations, the in-sample standard deviation for the sample covariance portfolio is equal to only 3.8%, while its out-of-sample counterpart is equal to 23.0%, more than 6 times larger. This provides further evidence to previous findings related to the large impact of estimation error, when using the sample covariance (e.g., Michaud, 1989). The other estimation techniques do a better job at minimizing the out-of-sample standard deviation and reducing the gap between in- and out-of-sample results. In particular, *tlasso* compares favourably to the alternative approaches, obtaining the lowest out-of-sample standard deviation in the FF 48 and S&P 100 daily datasets and performing well also in the FF 100 and S&P 100 monthly. Table 5 shows the differences between the out-of-sample standard deviation of *tlasso* portfolios and the ones optimized using other techniques. The confidence levels are computed using the Ledoit and Wolf bootstrap confidence interval for the ratio of two variances (Ledoit and Wolf, 2011). We observe that the out-of-sample standard deviation of *tlasso* is lower, and statistically significantly different than both *glasso* and LW in the FF 48 and S&P 100 daily datasets, while it is never higher and statistically significantly different from any other model.

We underline that *tlasso* in real-world scenarios shows better performance compared to *glasso*, while in the simulation study they obtain similar results.

This can be related to the robustness of *tlasso* to misspecification and outliers in the data (Finegold and Drton, 2011), and therefore the better capability of dealing with the typical fat-tail distribution of asset returns. The comparison between the results on the S&P 100 with daily and monthly returns allows to better characterize the relationship between *glasso* and *tlasso*. Indeed, from Table 5, we notice that the difference in the out-of-sample standard deviation of the *tlasso* and the *glasso* portfolios is statistically significant when using data with daily frequency, but not for the monthly ones. This may be due to the stylized property of *aggregational Gaussianity*, as reported by Cont (2001), which refers to the fact that the distribution of equity returns tends to have ticker tails for shorter time frequencies (e.g., daily), while being better approximated by a Gaussian distribution as the time frequency increases (e.g., monthly). Moreover, when the length of the estimation time interval is large compared to the number of asset weights to be estimated (e.g., 120 observations for FF 48), and consequently the precision matrices are less ill-conditioned, *tlasso* portfolios exhibit better out-of-sample risk properties than *glasso*. Finally, we notice that the equally weighted portfolios generally show high standard deviations compared to *glasso*, *tlasso*, RMT and LW. In two of the datasets, however, it performs better than the sample covariance matrix portfolios.

Concerning the Sharpe ratio, we obtain different results across the test cases: the sample covariance estimator shows the most inconsistent performance, with the highest out-of-sample Sharpe ratio in the FF 100 case (1.039) and the lowest in the FF 48 and S&P 100 monthly (0.384 and 0.151, respectively). The other estimators generally obtain good performance in all the test cases, with *glasso* and *tlasso* displaying the highest out-of-sample Sharpe ratios in the S&P 100 daily, RMT in the S&P 100 monthly and FF 48 and LW in the FF 100 case. Finally, we notice that the EW portfolio, despite beating the sample covariance portfolio in FF 48 and S&P100 monthly, does not seem to be competitive with the other methods in terms of risk adjusted performance.

Table 6 reports summary portfolio statistics. Columns 2 to 4 display the gross exposure (i.e., the sum of absolute values of the portfolio weights), the total negative exposures and the maximum negative exposure of individual assets. We see that the sample covariance portfolios are characterized by extreme exposures, especially for the FF 100 portfolios, where the gross exposures is more than 22 times higher than the initial endowment, due to a short exposure of 10.758. *glasso*, *tlasso*, RMT and LW show considerably

Table 4: In-sample and out-of-sample (oos) standard deviation, mean return and Sharpe ratio for for real-world data analysis.

	Standard deviation		Mean return		Sharpe ratio	
Model	in-sample	oos	in-sample	oos	in-sample	oos
Panel 1 - 48 Industry Portfolios (FF 48)						
sample cov.	0.068	0.124	0.070	0.048	1.042	0.384
glasso	0.083	0.110	0.074	0.077	0.889	0.697
tlasso	0.081	0.106	0.072	0.065	0.887	0.618
RMT	0.085	0.108	0.075	0.081	0.885	0.756
LW	0.074	0.113	0.068	0.070	0.917	0.620
EW	0.166	0.179	0.097	0.105	0.592	0.587
Panel 2 - 100 Size and Book-to-Market Portfolios (FF 100)						
sample cov.	0.038	0.230	0.228	0.239	6.147	1.039
glasso	0.085	0.127	0.102	0.069	1.189	0.546
tlasso	0.085	0.124	0.099	0.075	1.152	0.608
RMT	0.098	0.129	0.088	0.068	0.890	0.531
LW	0.070	0.121	0.105	0.092	1.496	0.758
EW	0.187	0.187	0.091	0.073	0.493	0.392
Panel 3 - S&P 100 2006-2016 - monthly data (S&P 100 mon.)						
sample cov.	0.039	0.206	0.100	0.031	2.652	0.151
glasso	0.084	0.119	0.099	0.079	1.211	0.663
tlasso	0.075	0.116	0.099	0.066	1.330	0.570
RMT	0.074	0.107	0.110	0.084	1.488	0.788
LW	0.069	0.115	0.094	0.052	1.362	0.453
EW	0.160	0.160	0.077	0.082	0.489	0.516
Panel 4 - S&P 100 2006-2016 - daily data (S&P 100 daily)						
sample cov.	0.067	0.138	0.090	0.090	1.531	0.649
glasso	0.086	0.124	0.110	0.108	1.411	0.872
tlasso	0.083	0.121	0.102	0.108	1.377	0.892
RMT	0.085	0.122	0.099	0.097	1.308	0.799
LW	0.074	0.125	0.100	0.094	1.475	0.755
EW	0.181	0.202	0.089	0.084	0.844	0.414

lower exposures, both in terms of whole portfolios and of individual securities. The EW portfolio, as it is long only by construction, is trivially the one with the lowest exposures. None of the methods promote sparsity of the

Table 5: Difference in out-of-sample standard deviations between the *tlasso* portfolios and the alternative methods. Negative numbers denote a lower standard deviation for *tlasso* compared to the alternative method (i.e., sample covariance, *glasso*, RMT, LW and EW). Statistical significance has been assessed with Ledoit and Wolf procedure (Ledoit and Wolf, 2011). *,**,*** denote 90%, 95% and 99% confidence level, respectively.

	sample cov.	glasso	RMT	LW	EW
FF 48	-0.019**	-0.004***	-0.002	-0.007*	-0.073***
FF 100	-0.106***	-0.002	-0.004	0.003	-0.063***
S&P 100 mon.	-0.091***	-0.003	0.009	0.001	-0.044***
S&P 100 daily	-0.018***	-0.003***	-0.001	-0.004***	-0.081***

weights, therefore in all cases the percentage of active position is 100%. The level of diversification is computed by the modified Herfindahl concentration coefficient H^*. Such measure can deal with short portfolio exposures and takes the lowest value for the most diversified portfolio (i.e., the EW portfolio). The concentration levels are similar for different portfolios, including the sample covariance one. This suggests that the main differences in the portfolio structure result from the allocation of weights and the level of gross exposures and not from the excessive concentration in a limited number of assets. Finally, we compute the turnover rate of the portfolios. The sample covariance portfolios show the worst performance in terms of turnover, due to both the amount of gross exposure and the estimation error. The equally weighted portfolio has zero turnover by construction and all the other techniques show considerably lower turnover levels than the sample covariance case.

5 Conclusion

The estimation of the precision matrix is fundamental to the implementation of investment strategies based on the minimum variance framework. In this paper, we consider two innovative methods based on Markovian graphs: *glasso* and *tlasso*. These techniques allow us to regularize the estimation of the precision matrix (i.e., the inverse of the covariance matrix) by imposing a constraint on the L_1-norm, assuming Gaussian and t-Student distributions, respectively. We test the models both on simulated and real world data,

measuring the quality of the estimation and the out-of-sample performances of the optimized portfolios. We compare them to the naive sample covariance estimator, equally weighted portfolios and two state-of-art techniques: random matrix and Ledoit Wolf shrinkage methods. According to our analysis, *glasso* and *tlasso* show interesting results: in the simulation framework they both improve the estimation of the precision matrix compared to the alternative techniques, reducing the bias and error of the estimates, and the actual risk in simulated portfolios, especially with ill-conditioned matrices. When applied to real data, they obtain good out-of-sample performances. *Tlasso*, which is more robust to misspecification and outliers, stands out for the low out-of-sample standard deviation, providing better results than *glasso* by just paying a small price in computational efficiency compared to *glasso*. The results are consistent across all the dataset considered, and the advantage of *tlasso* over *glasso* is larger when using data with daily frequency compared to monthly. Moreover, *glasso* and *tlasso* limit the portfolio short exposures and reduce considerably the turnover compared to the sample covariance estimator. High on the agenda, *glasso* and *tlasso* should be studied in other asset allocation frameworks. Indeed, the good portfolio performance, together with the simple implementation, make *glasso* and *tlasso* interesting tools for the Fintech industry and for the implementation of data-driven investment models, suitable also for distressed markets, when covariance matrices of the assets tend to be ill-conditioned.

A The *glasso* Algorithm

Here we briefly describe the algorithm proposed by Friedman et al. (2008) to solve (6), the *glasso* model. For convenience, we define X_i as the ith element of X, and and $X_\backslash i$ as the vector of all the elements of X except the ith. We also define the matrices $\mathbf{G}$ to be the estimate of $\mathbf{\Sigma}$, and $\mathbf{S}$ the sample covariance matrix. Furthermore, we identify the following partitions:[10]

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{\backslash i,\backslash i} & \mathbf{g}_{\backslash i,i} \\ \mathbf{g}'_{\backslash i,i} & g_{i,i} \end{pmatrix}, \qquad \mathbf{S} = \begin{pmatrix} \mathbf{S}_{\backslash i,\backslash i} & \mathbf{s}_{\backslash i,i} \\ \mathbf{s}'_{\backslash i,i} & s_{i,i} \end{pmatrix}. \tag{20}$$

Banerjee et al. (2008) show that the solution for $w_{\backslash i,i}$ can be computed by solving the following box-constrained quadratic program:

$$g_{\backslash i,i} = \arg\min_{y} \left\{ y' \mathbf{G}^{-1}_{\backslash i,\backslash i} y : ||y - \mathbf{s}_{\backslash i,i}||_\infty \leq \rho \right\}, \tag{21}$$

or in an equivalent way, by solving the dual problem

$$\min_{\beta^{(i)}} \left\{ \frac{1}{2} ||\mathbf{G}^{1/2}_{\backslash i,\backslash i} \beta^{(i)} - c||^2 + \rho ||\beta^{(i)}||_1 \right\}, \tag{22}$$

where $c = \mathbf{G}^{-1/2}_{\backslash i,\backslash i} \mathbf{s}_{\backslash i,i}$ and $\hat{\beta}^{(i)} = \mathbf{G}^{-1}_{\backslash i,\backslash i} \mathbf{g}_{\backslash i,i}$. As noted by Friedman et al. (2008), (22) resembles a lasso least square problem (see Tibshirani, 1996). The algorithm estimates then the ith variable on the others using as input $\mathbf{G}_{\backslash i,\backslash i}$, where $\mathbf{G}_{\backslash i,\backslash i}$ is the current estimate of the upper left block. The algorithm then updates the corresponding row and column of $\mathbf{G}$ using $\mathbf{g}_{\backslash i,i} = \mathbf{G}_{\backslash i,\backslash i} \hat{\beta}^{(i)}$ and cycles across the variables until convergence.

Glasso algorithm

1. Start with $\mathbf{G} = \mathbf{S} + \rho \mathbf{I}$. The diagonal of $\mathbf{G}$ is unchanged in the next steps.

2. For each $i = 1, 2, \ldots, n, 1, 2, \ldots, n, \ldots$, solve the lasso problem (22), which takes as input $\mathbf{G}_{\backslash i,\backslash i}$ and $\mathbf{s}_{\backslash i,i}$. This gives a $n-1$ vector solution $\hat{\beta}$. Fill in the corresponding row and column of $\mathbf{G}$ using $\mathbf{g}_{\backslash i,i} = \mathbf{G}_{\backslash i,\backslash i} \hat{\beta}$.

3. Repeat until a convergence criterion is satisfied.

[10] The dimension of $\mathbf{G}_{\backslash i,\backslash i}$, $g_{\backslash i,i}$ and $g_{i,i}$ are respectively $((n-1) \times (n-1))$, $((n-1) \times 1)$ and (1×1).

The algorithm has a computational complexity of $O(n^3)$ for dense problems, and considerably less than that for sparse problems (Friedman et al., 2008).

B Alternative Covariance Estimation Methods

Here, we briefly describe the benchmark covariance estimators we use in the comparative analysis. Differently from *glasso* and *tlasso*, these approaches provide an estimate for the covariance matrix and not for the precision matrix. Hence, we compute the precision matrix for such methods to be plug-in into the minimum variance portfolio by inverting the covariance.

In particular, we consider the *sample covariance* and the equally weighted methods (that are commonly regarded as naive approaches) and two state-of-art estimators: random matrix theory and Ledoit Wolf Shrinkage.

The equally weighted (EW) portfolio, a tough benchmark to beat (DeMiguel et al., 2009b), can be interpreted as an extreme shrinkage estimator of the global minimum variance portfolio, obtained using the identity matrix as the estimate of the covariance matrix. Indeed, using (3), we obtain $\hat{\mathbf{w}}_{EW} = \frac{\mathbf{I1}}{\mathbf{1'I1}} = \frac{1}{n}\mathbf{1}$. By assuming zero correlations and equal variances, such approach is very conservative in terms of estimation error and it suitable in case of severe unpredictability of the parameters.

The second naive approach is the sample covariance estimator, defined as:

$$\mathbf{S} = \frac{1}{t-1}\sum_{\tau=1}^{t}(X_\tau - \bar{X})(X_\tau - \bar{X})', \tag{23}$$

where t is the length of the estimation period, X_i is the multivariate variate vector of assets' returns at time τ and $\bar{X}$ is the vector of the average return for the n assets. Such estimator, when computed on datasets with a number of asset close to the length of the window size, is typically characterized by a larger eigenvalue dispersion compared to true covariance matrix, causing the matrix to be ill-conditioned (Meucci, 2009). Therefore, when computing the precision matrix by inverting the covariance matrix, estimates are typically

not reliable and unstable on different samples as its ill-conditioning nature *amplifies* the effects of the estimation error in the covariance matrix.

The shrinkage methodology of Ledoit-Wolf (LW) is well-known to better control for the presence of estimation errors, especially for datasets with a large ratio of n/t, where n is the number of assets and t the length of the estimation window. The Ledoit-Wolf shrinkage estimator is defined to be a convex combination of the sample covariance matrix $\mathbf{S}$ and $\widehat{\mathbf{\Sigma}}_T$, a highly structured target estimator, such that $\widehat{\mathbf{\Sigma}}_{LW} = a\mathbf{S} + (1-a)\widehat{\mathbf{\Sigma}}_T$ with $a \in [0,1]$. Following Ledoit and Wolf (2004a), we consider as structured estimator $\widehat{\mathbf{\Sigma}}_T$ the constant correlation matrix, such that all the pairwise correlations are identical and equal to the average of all the sample pairwise correlations. As the target estimator is characterized by good conditioning, the resulting shrinkage estimator $\widehat{\mathbf{\Sigma}}_{LW}$ has a smaller eigenvalues dispersion than the sample covariance matrix. In fact, the sample covariance matrix is *shrunk* towards the structured estimator, with intensity depending on the value of the *shrinkage constant* a. Ledoit-Wolf estimation of a is based on the minimization of the expected distance between $\widehat{\mathbf{\Sigma}}_{LW}$ and $\mathbf{\Sigma}$. For further details, the reader is referred to Ledoit and Wolf (2004a).[11]

The last approach we focus on is the so called random matrix theory (RMT) estimator $\widehat{\mathbf{\Sigma}}_{RMT}$, introduced by Laloux et al. (1999). The approach is based on the fact that, in the case of financial time series, the smallest eigenvalues of the correlation matrices are often dominated by noise. From the known distribution of the eigenvalues of a random matrix, it is possible then to filter out the part of spectrum that is likely associated with estimation error and maintain only the eigenvalues that carry useful information (Laloux et al., 1999). In particular, when assuming *i.i.d.* returns, the eigenvalues of the sample correlation matrix are then distributed according to a Marcenko-Pastur (MP) distribution as a consequence of the estimation error. Therefore, we can compute the eigenvalues that correspond to noise based on the minimum

[11] Interestingly, the Ledoit-Wolf shrinkage is closely related to portfolio optimization with L_2 penalization of weight estimates. Indeed, the optimization problem $\min_{\mathbf{w}\in C}(\mathbf{w}'\widehat{\mathbf{\Sigma}}\mathbf{w} + a\mathbf{w}'\mathbf{w})$, with $C = \{\mathbf{w}|\mathbf{1}'\mathbf{w} = 1\}$ can be equivalently stated as $\min_{\mathbf{w}\in C}(\mathbf{w}'(\widehat{\mathbf{\Sigma}} + a\mathbf{I})\mathbf{w})$, which then is equivalent to solving the problem using the Ledoit-Wolf shrinkage estimator with $\widehat{\mathbf{\Sigma}}_T = \mathbf{I}$ (Bruder et al., 2013).

and maximum eigenvalues of the theoretical distribution, such that:

$$\lambda_{\min \max} = \sigma^2\big(1 \pm \sqrt{n/t}\big)^2, \tag{24}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the theoretical smallest and largest eigenvalues in a $n \times n$ random covariance matrix estimated by a sample of t observations and σ^2 is the variance of the i.i.d. asset returns. Only the eigenvalues outside the interval $[\lambda_{\min}, \lambda_{\max}]$ are then assumed to bring useful information, while the others correspond to noise. Here, we estimate the covariance matrix then by *eigenvalue clipping*, a technique that consists in substituting the eigenvalues smaller than $\lambda_{\max}$ with their average:

$$\widehat{\mathbf{\Sigma}}_{RMT} = \mathbf{V}\mathbf{\Lambda}_{RMT}\mathbf{V}', \tag{25}$$

where $\mathbf{V}$ represents the eigenvectors of the sample covariance matrix and $\mathbf{\Lambda}_{RMT}$ is the diagonal matrix with the ordered eigenvalues, where the eigenvalues $\lambda \leq \lambda_{\max}$ are substituted by their average (Bouchaud and Potters, 2009). The RMT filtering has then the effect of averaging the lowest eigenvalues, improving the conditioning of the matrix and therefore reducing the sensitivity of the precision matrix to estimation errors.

For further details the reader is refereed to Laloux et al. (1999), Bouchaud and Potters (2009) and Bruder et al. (2013).

Acknowledgement

Sandra Paterlini acknowledges ICT COST Action IC1408 from CRoNoS. Gabriele Torri acknowledges the support of the Czech Science Foundation (GACR) under project 15-23699S and SP2017/32, an SGS research project of VSB-TU Ostrava. Rosella Giacometti and Gabriele Torri acknowledge the support given by University of Bergamo research funds 2016 2017.

References

Baba K, Shibata R, Sibuya M (2004) Partial correlation and conditional correlation as measures of conditional independence. Australian & New Zealand Journal of Statistics 46(4):657–664

Banerjee O, Ghaoui LE, d'Aspremont A (2008) Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. Journal of Machine Learning Research 9:485–516

Black F, Litterman R (1992) Global portfolio optimization. Financial Analysts Journal 48(5):28–43

Bouchaud JP, Potters M (2009) Financial applications of random matrix theory: a short review. arXiv preprint arXiv:09101205

Brodie J, Daubechies I, De Mol C, Giannone D, Loris I (2009) Sparse and stable Markowitz portfolios. Proceedings of the National Academy of Science 106(30):12,267–12,272

Brownlees CT, Nualart E, Sun Y (2015) Realized networks. Working Paper, available on SSRN

Bruder B, Gaussel N, Richard JC, Roncalli T (2013) Regularization of portfolio allocation. Working Paper, available on SSRN

Cont R (2001) Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1:223–236

DeMiguel V, Nogales FJ (2009) Portfolio selection with robust estimation. Operations Research 57:560–577

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

DeMiguel V, Garlappi L, Uppal R (2009b) Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? Review of Financial Studies 22(5):1915–1953

Dempster AP (1972) Covariance selection. Biometrics 28(1):157–175

Engle R (2002) Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics 20(3):339–350

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

Finegold M, Drton M (2011) Robust graphical modeling of gene networks using classical and alternative t-distributions. The Annals of Applied Statistics 5(2A):1057–1080

Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3):432–441

Friedman J, Hastie T, Tibshirani R (2014) glasso: Graphical lasso-estimation of gaussian graphical models. R package

Goto S, Xu Y (2015) Improving mean variance optimization through sparse hedging restrictions. Journal of Financial and Quantitative Analysis 50(6):1415–1441

Højsgaard S, Edwards D, Lauritzen S (2012) Graphical models with R. Springer Science & Business Media

Kan R, Zhou G (2007) Optimal portfolio choice with parameter uncertainty. Journal of Financial and Quantitative Analysis 42(3):621–656

Kolm PN, Tütüncü R, Fabozzi F (2014) 60 years following harry markowitz's contribution to portfolio theory and operations research. European Journal of Operational Research 234 (2):343–582

Kotz S, Nadarajah S (2004) Multivariate t-distributions and their applications. Cambridge University Press

Kremer PJ, Talmaciu A, Paterlini S (2017) Risk minimization in multi-factor portfolios: What is the best strategy? Annals of Operations Research pp 1–37

Laloux L, Cizeau P, Bouchaud JP, Potters M (1999) Noise dressing of financial correlation matrices. Physical Review Letters 83(7):1467–1469

Lam C, Fan J (2009) Sparsistency and rates of convergence in large covariance matrix estimation. Annals of Statistics 37(6B):4254

Lange KL, Little RJ, Taylor JM (1989) Robust statistical modeling using the t distribution. Journal of the American Statistical Association 84(408):881–896

Lauritzen SL (1996) Graphical models, vol 17. Clarendon Press

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

Ledoit O, Wolf M (2004b) A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis 88(2):365–411

Ledoit O, Wolf M (2011) Robust performances hypothesis testing with the variance. Wilmott 2011(55):86–89

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

McLachlan GJ, Krishnan T (1997) The EM Algorithm and Extensions, Second Edition. Wiley Online Library

Meucci A (2009) Risk and asset allocation. Springer Science & Business Media

Michaud RO (1989) The Markowitz optimization enigma: Is optimized optimal? ICFA Continuing Education Series 1989(4):43–54

Murphy KP (2012) Machine learning: a probabilistic perspective. The MIT Press

Rothman AJ, Bickel PJ, Levina E, Zhu J, et al. (2008) Sparse permutation invariant covariance estimation. Electronic Journal of Statistics 2:494–515

Stevens GV (1998) On the inverse of the covariance matrix in portfolio analysis. The Journal of Finance 53(5):1821–1827

Tibshirani R (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological) 58(1):267–288

Witten DM, Friedman JH, Simon N (2011) New insights and faster computations for the graphical lasso. Journal of Computational and Graphical Statistics 20(4):892–900

Won JH, Lim J, Kim SJ, Rajaratnam B (2013) Condition-number-regularized covariance estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75(3):427–450

Yuan M, Lin Y (2007) Model selection and estimation in the Gaussian graphical model. Biometrika 94(1):19–35

Table 6: Portfolio statistics on real-world data. The Table reports, from left to right, gross exposure (gross exp.) ($\sum_i |w_i|$), total short exposure (short exp), maximum negative exposure of individual assets (max short), percentage of active positions in the portfolios (active pos.), modified Herfindahl diversification index corrected to account for short portfolio ($H^* = \sum_i w_i'^2$ where $w_i' = w_i/(\sum_i |w_i|)$) and portfolio turnover (turnover). The reported values are the average across all the rebalancing periods.

Model	gross exp.	short exp.	max short	active pos.	H^*	turnover
Panel 1 - 48 Industry Portfolios						
sample cov.	4.426	1.713	-0.213	100%	0.036	1.149
glasso	2.142	0.571	-0.068	100%	0.040	0.282
tlasso	2.253	0.627	-0.070	100%	0.039	0.317
RMT	1.894	0.447	-0.046	100%	0.046	0.201
LW	2.866	0.933	-0.103	100%	0.040	0.455
EW	1.000	0.000	0.000	100%	0.021	0.000
Panel 2 - 100 Size and Book-to-Market Portfolios						
sample cov.	22.516	10.758	-0.807	100%	0.017	11.799
glasso	4.656	1.828	-0.121	100%	0.018	0.703
tlasso	4.573	1.786	-0.114	100%	0.018	0.798
RMT	3.022	1.011	-0.062	100%	0.022	0.272
LW	5.512	2.256	-0.141	100%	0.018	0.973
EW	1.000	0.000	0.000	100%	0.010	0.000
Panel 3 - S&P 100 2006-2016 (monthly data)						
sample cov.	6.954	2.977	-0.297	100%	0.020	3.016
glasso	1.488	0.244	-0.025	100%	0.022	0.309
tlasso	1.707	0.353	-0.030	100%	0.025	0.271
RMT	1.850	0.425	-0.036	100%	0.023	0.258
LW	2.190	0.595	-0.039	100%	0.024	0.338
EW	1.000	0.000	0.000	100%	0.011	0.000
Panel 4 - S&P 100 2006-2016 (daily data)						
sample cov.	4.569	1.785	-0.186	100%	0.022	3.192
glasso	2.376	0.688	-0.061	100%	0.023	1.058
tlasso	2.479	0.739	-0.058	100%	0.025	1.113
RMT	2.121	0.561	-0.043	100%	0.026	0.888
LW	3.043	1.021	-0.078	100%	0.025	1.650
EW	1.000	0.000	0.000	100%	0.011	0.000