Sparse Estimation of Precision Matrix in Portfolio Optimization

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Author Name Jari Nelisse

Abstract

The precision matrix is important in portfolio risk minimization and consists of hedge trades in which a stock is hedged by all the other stocks in the portfolio. However, in practice with finite samples, estimation of the precision matrix is subject to multicollinearity, which makes the hedge trades too unstable and unreliable. Therefore, we propose two shrinkage estimators with the aim to reduce the number of stocks and shrink trade sizes in each hedge trade, to obtain a "sparse" estimator of the precision matrix. The proposed estimators compare favorably with respect to other methods (equal weighting, shrunk covariance matrix, nonnegativity constraints) and achieve significant out-of-sample risk reductions.

Bachelor Thesis

Econometrie en Operationele Research

Student ID: 445883

Supervisor: dr. (Rasmus) SOR Lonn

Date: 04-07-2021

Contents

1 Introduction

In the Modern Portfolio Theory, founded by Nobel laureate Markowitz (1952), the precision matrix of the stock returns, Σ^{-1}, plays a crucial role regarding mean variance portfolio optimization. Essentially, optimal portfolio weights w can be obtained by the following proportionality relationship $w \propto \Sigma^{-1}\mu$, where μ are the expected returns of the assets in the portfolio. Therefore, we propose two versions of a mean variance optimizer for the precision matrix, where the mean variance optimizer is generally defined as $\Psi \equiv \Sigma^{-1}$.

When the objective is to find a set of weights w that yield the lowest portfolio variance possible, we use the mean variance optimizer to obtain the global minimum variance (GMV) portfolio $w \propto \Sigma^{-1}\mathbf{1}_N$. In this paper we use two variants of the mean variance optimizer in order to achieve risk reduction, which will be introduced later.

Although the Modern Portfolio Theory provides a clear and strong theoretical foundation, empirical results of J David Jobson and R. M. Korkie (1981) and DeMiguel, Garlappi, and Uppal (2009) show that mean variance optimization is difficult to implement in practice. The sample covariance matrix seems to play an important role in this because it tends to be susceptible to large estimation errors. Such errors have a tendency to occur when there are not enough historical stock return observations (T) relative to the amount of stocks in the portfolio (N) (i.e. larger N/T ratio), or when the correlations between the stocks in the portfolio are high. Kan and Zhou (2007) found that when N is relatively large compared to T, estimation errors in the sample covariance matrix are the most problematic. Empirically, this poor estimation lead to unstable and unreliable optimal weights.

This paper builds upon the work of Goto and Xu (2015). The aim is to improve the estimation of the precision matrix, i.e. the mean variance optimizer, $\Psi_l \equiv \Sigma^{-1}$, $l \in \{\rho, \lambda\}$. Here Ψ_ρ denotes the proposed mean variance optimizer of Goto and Xu (2015), and Ψ_λ is our own proposed mean variance optimizer. Stevens (1998) have found that Σ^{-1} possesses optimal hedging relations in their rows (or columns), that is, the row of stock ith is proportional to that stock's hedge portfolio. In this hedge portfolio, a long position is taken in stock i and a short position in the other $N-1$ stocks. Thereby, the $N-1$ stocks track the ith stock return to minimize the tracking error variance. This implies that the off-diagonal elements of row $i, i \in \{1, ..., N\}$ in ψ_p, can be regarded as regression coefficients from a scaled ordinary least squares (OLS) estimation of stock i on the other $N-1$ stocks. The link of the elements of Ψ_k with a regression model opens the door for more advanced Econometric methods to reduce estimation errors and improve robustness.

In a general regression model, when the number of explanatory variables (N) are high relative to the number of observations (T), multicollinearity subjects the coefficients to large estimation errors. The same sort of problem also arises in the context of portfolio optimization. To conquer this multicollinearity problem and make the proposed mean variance optimizer sparse (i.e. set some off-diagonal elements to 0), we propose two strategies. The strategies are based on penalization procedures, where in every row i of ψ the regression coefficients are shrunk toward 0 to curtail extreme hedge positions, and redundant stocks are dropped in each hedge portfolio (subset selection).

The first strategy is the one of Goto and Xu (2015). To obtain shrinkage and sparsity when estimating ψ_ρ, the 'lasso' method by R. Tibshirani (1996) served as an important tool. With this method, the l_1 norm, i.e. the sum of the absolute values of the off-diagonal elements of Ψ_ρ (which serves as an additional regularizer penalty in the cost function), is penalized to tackle multicollinearity. Because the precision matrix needs to be positive definite and symmetric, Goto and Xu (2015) used quasi-maximum likelihood (QML) with a contraint on the l_1 norm, similarly to Yuan and Lin (2007) and Rothman et al. (2008). This way the N hedge regressions are estimated jointly as a group. Friedman, Hastie, and R. Tibshirani (2008) have found that the QML method is equivalent to a N-coupled lasso problem. So the 'graphical lasso' (glasso) algorithm can be employed, as also done by Goto and Xu (2015), to solve the QML estimation problem.

The second strategy is based on the approach of Zou and Hastie (2005), namely Elastic Net regression. This yields an Elastic Net type of penalty with a combination of the l_1 and l_2 norm, (where the l_2 norm is in this case the sum of the squared values of the off-diagonal elements of Ψ_λ). The algorithm builds upon the glasso algorithm of Friedman, Hastie, and R. Tibshirani (2008). This approach also uses the same hedging relations to achieve shrinkage and subset selection. However, it also includes a diagonal target matrix for the precision matrix, in the form of an identity matrix.

The main goal of the two mean variance optimizers, Ψ_ρ and Ψ_λ, is to achieve risk reduction. To assess their out-of-sample performance, we used the first three data sets also used by Goto and Xu (2015). This data is publicly available on the Web site of Ken French (http://mba.tuck .dartmouth.edu/pages/faculty/ken.french/data library.html). The portfolios used for evaluation are based on size and book-to-market ratio, as well as Fama and French's (1997) 48 industry portfolios. N/T ratios range from 0.400 to 1.233. To look whether risk reduction is achieved with our proposed estimators Ψ_ρ and Ψ_λ, we compare GMV portfolio performances of our proposed estimators with different estimators for Σ^{-1}.

When we assess the portfolios empirically, we see that Ψ_ρ and Ψ_λ both achieve significant out-of-sample portfolio risk reduction compared to the alternative portfolios. Thereby, the Sharpe ratios of our proposed mean variance optimizers are both substantially high and significantly higher compared to the Sharpe ratios of much of our alternative portfolios. Lastly, when we look at the certainty equivalent returns adjusted for turnover/transaction costs, the results of the portfolios with Ψ_ρ and Ψ_λ are promising.

The rest of the paper has the following organisation. Section 2 is the theory section, where we give a more detailed explanation about the hedging relations in Σ^{-1} founded by Stevens (1998), as well as a literature review about the subject and our proposed estimators. Section 3 is the data section which shows how the data has been gathered and processed. In section 4 the methodology section can be found, where the used methods are presented. Section 6 contains the results of this paper. Finally, we finish the paper with the conclusion in section 7.

2 Theory

2.1 Hedging Relations In The Precision Matrix

Stevens (1998) found out that the precision matrix contains optimal hedge relations of the stocks. Specifically, the rows (or columns) consist of proportional hedge portfolios, where in the hedge portfolio for stock i a long position is taken in stock i and a short position in the other $N-1$ stocks. This short position tracks the return of stock i in order to minimize the variance of the tracking error without any constraints. The following regression formula shows how the hedging portfolio can be estimated:

$$r_{i,t} = \alpha_i + \sum_{k=1,k\neq i}^{N} \beta_{i|k} r_{k,t} + \epsilon_{i,t}, \tag{1}$$

where $r_{i,t}$ is the return for stock i in period t; $\beta_{i|k}$ is the marginal contribution of stock k to the hedge of stock i beyond the other $N-2$ stocks in the hedging portfolio; and $\epsilon_{i,t}$ is the idiosyncratic component of stock i, where the variance of $\epsilon_{i,t}$ is denoted by $v_i = var(\epsilon_{i,t})$, i.e. the idiosyncratic risk of stock i. In each hedge regression, the objective is to minimize v_i. Therefore, the regression in (1) can be regarded as an OLS estimation problem.

Stevens (1998) calls the formula in (1) a "regression hedge" and with the help of this regression equation the hedging relations in the precision matrix can be made clear. That is, if we have an N x N precision matrix denoted by $\Sigma^{-1} = \Psi = [\psi_{ij}]$ (where ψ_{ij} denotes the (i,j)th element of Ψ), the following relationship between (1) and Ψ can be obtained (as founded by Stevens (1998)):

$$\psi_{ij} = \begin{cases} -\frac{\beta_{i|j}}{v_i} & \text{if i} \neq \text{j} \\ \frac{1}{v_i} & \text{if i = j} \end{cases} \tag{2}$$

As already discussed, the $\beta_{i|j}$ coefficient represents a marginal hedging ability for stock j to hedge stock i beyond the other $N-2$ stocks. Therefore, ψ_{ij} can also be viewed as a measurement, scaled by v_i, of marginal hedging ability between stocks i and j conditional on all the other stocks in the concerned portfolio. The relationship shown by (2) also reveals that if stocks i and j have no correlation with each other, conditional on the other stocks in the portfolio, $\beta_{i|j} = 0$, which implies that $\psi_{ij} = 0$.

To get a more intuitive feeling about this important relationship, the ith row of the precision matrix can be represented as follows:

$$\Psi(i, \cdot) = \frac{1}{v_i}[-\beta_{i|1}, ..., -\beta_{i|i-1}, 1, -\beta_{i|i+1}, ..., -\beta_{i|N}]. \tag{3}$$

This equation shows that the ith row of the precision matrix can be regarded as a vector of stock holdings in the ith stock's hedge portfolio. The coefficients imply that a unit long position is taken in stock i (the 1 coefficient), and a short position in the hedge portfolio denoted by the regression in equation (1). The holdings of each stock are also scaled by $\frac{1}{v_i}$, which makes that when the unhedgeable risk of stock i is smaller, the optimal portfolio takes a larger position in that particular stock. So in order to minimize portfolio risk, we have shown that hedge trades play an important role when working with the mean variance optimizer Ψ.

2.2 Improved Estimation Of The Precision Matrix

2.2.1 Dealing With Multicollinearity

Estimation of the precision matrix Σ^{-1} plays a crucial role in mean variance portfolio optimization. With the described characteristics of Σ^{-1} in the preceding section, it is clear that mean variance optimization is prone to large estimation errors. A hedge regression, as shown in equation (1), contains a constant and $N-1$ stocks as independent variables. In many cases these stock returns are highly correlated. Also in practice, the number of available historical returns relative to the number of stocks to estimate the regression is not sufficient enough to get a reliable estimation result. Therefore, in most practical situations the hedge regression estimation suffers from multicollinearity. The consequences of multicollinearity are that the estimated hedge coefficients ($\hat{\beta}'s$) are inefficient, i.e. have large estimation errors. This gives unstable and unreliable estimates of the ($\hat{\beta}'s$). This implies that the off-diagonal elements of Ψ_i are also prone to large estimation errors, as made clear by the relationships given in equation (2).

To handle the multicollinearity problem, one first solution might be to get higher frequency returns, such as daily returns, to estimate the sample covariance matrix. Jagannathan and Ma (2003) used such daily returns instead of monthly for estimating the sample covariance matrix and found more reliable estimates. However, daily returns are less used in the day to day practice by investment professionals and monthly returns are more of a standard. So our approach to conquer the multicollinearity problem is to penalize the the l_1 and l_2 norms of the regression coefficients that need to be estimated.

2.2.2 Estimation With Glasso Estimator

With the mean variance optimizer proposed by Goto and Xu (2015), Ψ_ρ, the l_1 norm is penalized to tackle the multicollinearity problem. R. Tibshirani (1996) proposed the "lasso" method, where the sum of the absolute values of the regression coefficients are penalized with this l_1 norm. When we connect this with the hedge regression in equation (1), we obtain the following "lasso" estimation problem:

$$\hat{\beta}^{lasso}_{i|k} = argmin_\beta \left\{ \sum_{t=1}^{T} (r_{i,t} - \sum_{k=1,k\neq i}^{N} \beta_{i|k} r_{k,t})^2 + \gamma \sum_{k=1,k\neq i}^{N} |\beta_{i|k}| \right\}. \tag{4}$$

When the regressors are orthonormal, we obtain the following relationship between the lasso coefficient $\hat{\beta}^{lasso}_{i|k}$ and the OLS coefficient $\hat{\beta}^{OLS}_{i|k}$:

$$\hat{\beta}^{lasso}_{i|k} = (|\hat{\beta}^{OLS}_{i|k}| - \gamma/2)_+ sign(\hat{\beta}^{OLS}_{i|k}); k = 1, ..., N, k \neq i, \tag{5}$$

where $(x)_+ = max(x, 0)$, and the penalty parameter γ is the soft threshold. Absolute values of OLS point estimates below this soft threshold give a lasso solution $\hat{\beta}^{lasso}_{i|k}$ which is set to 0. When the absolute value is above the soft threshold, $\hat{\beta}^{lasso}_{i|k}$ is shrunk toward 0 by the magnitude of the soft threshold, but it never actually crosses 0 or alternates sign. With this approach the lasso solution achieves shrinkage and subset selection in each hedge regression.

However, the orthonormal condition necessary for the relationship in (5) to hold is hardly satisfied, and we often see correlations between the regressors. When this is the case, the $\hat{\beta}^{lasso}_{i|k}$ can be obtained in an iterative way. Denote with $\tilde{\beta}^{(\gamma)}_{i|k}$ the current estimate for $\beta_{i|k}$ at penalty parameter γ. Then, Friedman, Hastie, Höfling, et al. (2007) find the unique convergence to the following lasso estimate:

$$\tilde{\beta}_{i|k}^{(\gamma)} = \boldsymbol{S}(\tilde{\beta}_{i|k}^{(\gamma)} + \sum_{t=1}^{T} \left\{ r_{k,t} x(r_{i,t} - \sum_{j=1, j \neq i}^{N} \tilde{\beta}_{i|k}^{(\gamma)} r_{j,t}) \right\}, \gamma); k = 1, ..., N, k \neq i, \tag{6}$$

where $\boldsymbol{S}(b, \gamma) = sign(b)(|b| - \gamma)_+$ denotes the soft-thresholding operator.

When the number of assets is larger than the number of return observations for the assets, i.e. $N > T$, each hedge regression has an infinite number of least-squares solutions. However, the lasso algorithm is still able to obtain an unique solution, as long as the regressors are continuous random variables, regardless what the specific numbers of N and T are (R. J. Tibshirani et al. (2013)).

The value from which we start the iteration can be obtained with different methods, because the starting value has no influence on the final solution. In practice, the Moore–Penrose pseudoinverse (via singular value decomposition) can be used to provide us the starting iteration for the least-squares estimate.

Because of the hedging relations which Stevens (1998) found in the precision matrix, we could apply lasso to each hedge regression to estimate each row (or column) in order to achieve shrinkage and variable selection. Meinshausen, Bühlmann, et al. (2006) already implemented such a method, which they called nodewise-regression, and found asymptotically correct estimates of the nonzero elements under certain conditions (in more general context). Callot et al. (2019) used a similar approach, but then in the context of portfolio optimization, and also found promising results compared with other shrinkage based approaches. However, the precision matrix needs to be positive definite and symmetric for portfolio optimization, and the proposed row-by-row (or column-by-column) lasso estimation (nodewise-regression) does not restrict the precision matrix towards those conditions. This could be a problem in practical applications, especially now in our portfolio optimization method. Therefore, a joint instead of separate estimation of the N hedge regressions can help to prevent this problem.

To obtain a joint estimation, i.e. all elements of the precision matrix are estimated all at once, we follow the methods used by Yuan and Lin (2007), Banerjee, El Ghaoui, and d'Aspremont (2008), and Friedman, Hastie, and R. Tibshirani (2008). They use a quasi-maximum likelihood (QML) method with l_1 norm penalty of its off-diagonal elements. The QML estimation problem (7) can be solved with the graphical lasso (glasso) algorithm of Friedman, Hastie, and R. Tibshirani (2008), because they demonstrated that this problem (7) is equivalent to a N-coupled lasso problem. With the glasso algorithm the following negative l_1-regularized likelihood function is minimized:

$$argmin_{\Psi_\rho=[\psi_{ij}]_\rho}\left\{-ln(det(\Psi_\rho))+trace(\hat{S}\Psi_\rho)+\rho||\Psi_\rho||_1\right\}, \tag{7}$$

where det and $trace$ denote respectively the determinant and trace of the matrix, $\hat{S}$ is the sample covariance matrix, ρ is the regularization parameter and $||\Psi_\rho||_1$ denotes the l_1 norm of the off-diagonal elements of Ψ_ρ, where the off-diagonal elements can be specified as $\sum_{i=1,i\neq j}^{N}\sum_{j=1,j\neq i}^{N}|\psi_{ij}|_\rho$. The value of $\rho\geq 0$ denotes the penalty given to the the l_1 norm of the off-diagonal elements of The larger the value of ρ, the more the sparsity of Ψ_ρ is promoted. Note that when $\rho=0$, the last term of the formula in (7) falls away, so then the solution becomes the same as the unconstrained QML solution. The optimal choice of ρ will be discussed in the methodology section.

2.2.3 Estimation With Elastic Net Estimator

The other mean variance optimizer we propose is Ψ_λ and here the l_1 and l_2 norms are penalized to conquer multicollinearity. Lasso regression is a regression tool which uses the l_1 norm, and ridge regression uses the l_2 norm. To achieve a combination of them we use the elastic net regression method of Zou and Hastie (2005). The estimation problem can be denoted as follows:

$$\hat{\beta}_{i|k}^{elastic}=argmin_\beta\left\{\sum_{t=1}^{T}(r_{i,t}-\sum_{k=1,k\neq i}^{N}\beta_{i|k}r_{k,t})^2+\gamma\sum_{k=1,k\neq i}^{N}|\beta_{i|k}|+\eta\sum_{k=1,k\neq i}^{N}\beta_{i|k}^2\right\}. \tag{8}$$

Note that compared to formula (4), we now have two penalty parameters γ and η. In addition to the l_1 norm now the l_2 norm is also added, where the sum of the squared values of the regression coefficients are penalized with this norm. With an orthogonal design, the following relationship between $\hat{\beta}_{i|k}^{elastic}$ and $\hat{\beta}_{i|k}^{OLS}$ can be defined:

$$\hat{\beta}_{i|k}^{elastic}=\frac{(|\hat{\beta}_{i|k}^{OLS}|-\gamma/2)_+}{1+\eta}sign(\hat{\beta}_{i|k}^{OLS});k=1,...,N,k\neq i. \tag{9}$$

Because the regressors are hardly orthonormal, the relationship shown in (9) does not hold in general. Thus, to obtain an elastic net estimation for the precision matrix, we use a modified version of the glasso algorithm of proposed by Kovács et al. (2021). This gives rise to the following negative likelihood which needs to be minimized:

$$argmin_{\Psi_\rho=[\psi_{ij}]_\rho}\left\{-ln(det(\Psi_\rho))+trace(\hat{S}\Psi_\rho)+\lambda(\alpha||\Psi_\rho||_1+\frac{1-\alpha}{2}||\Psi_\rho||_2)\right\}, \tag{10}$$

where λ is a regularization parameter, $\alpha\in[0,1]$ is the tuning parameter and $\hat{S}$ is the sample covariance matrix. Again we have the l_1 norm of the off-diagonal elements of Ψ_ρ denoted by $||\Psi_\rho||_1$, with off-diagonal elements specified as $\sum_{i=1,i\neq j}^{N}\sum_{j=1,j\neq i}^{N}|\psi_{ij}|_\rho$. Now we also have the l_2 norm of the off-diagonal elements of Ψ_ρ, denoted by $||\Psi_\rho||_2$. Here the off-diagonal elements are specified as $\sum_{i=1,i\neq j}^{N}\sum_{j=1,j\neq i}^{N}(\psi_{ij})_\rho^2$. Minimizing the expression in (10) is called the Graphical Elastic Net (gelnet) (Kovács et al. (2021)).

Note that instead of two regularization parameters for both the l_1 and l_2 norm, we only have one regularization parameter λ. The tuning parameter α can be set to values between 0 and 1, where setting α to 0 gives us the ridge regression and setting α to 1 gives us the lasso regression. For the elastic net penalty, α is set to be 0.5.

In the gelnet algorithm we can use different target matrices. With a target matrix included, the estimation problem now becomes:

$$argmin_{\Psi_\rho=[\psi_{ij}]_\rho}\left\{-ln(det(\Psi_\rho))+trace(\hat{S}\Psi_\rho)+\lambda(\alpha||\Psi_\rho-T||_1+\frac{1-\alpha}{2}||\Psi_\rho-T||_2)\right\}. \quad (11)$$

This expression also contains a known diagonal positive definite target matrix T. The inclusion of such target matrices which are suitable in the context could considerably improve estimation results. An arbitrary positive semi-definite matrix might be more flexible than a diagonal matrix, but in practice many target matrices used for this purpose were diagonal matrices. For example, all target matrices that Van Wieringen and Peeters (2016) and Kuismin, Kemppainen, and Sillanpää (2017) used were diagonal matrices. The target matrix we use in our gelnet procedure is the identity matrix, because we consider this target as a good way to promote sparsity.

3 Data

To evaluate the out-of-sample performance of our proposed mean variance optimizers $\hat{\Psi}_\rho$ and $\hat{\Psi}_\lambda$, we use the first three data sets also used by Goto and Xu (2015). In Table 1 some characteristics about the data are presented. The data sets are all from assets in U.S markets. All the data sets have monthly returns ranging from July 1963 to December 2010. The data has been gathered on Web site of Ken French (http://mba.tuck.dartmouth.edu/ pages/faculty/ken.french/data library.html).

The first data set consists of 100 portfolios formed on size and book-to-market ratio (data set 1), the second data set of Fama and French's (1997) 48 industry portfolios (data set 2), and the third one a combination of the preceding two data sets consisting of 148 portfolios (data set

3).

All the returns are exclusive dividend, and the risk-free rate is subtracted from all the returns to obtain excess returns. When a particular return was missing, excess market return data from the Fama/French 5 Factors model with compatible date from the same Web site of Ken French replaced that particular return.

As also mentioned by Goto and Xu (2015), the data sets used have relatively many assets compared to the number of available observations per asset, which causes higher N/T ratios. We already discussed that data with high N/T ratios are more prone to estimation errors, so our data gives a good opportunity to test how $\hat{\Psi}_\rho$ handles in an environment where such errors are more likely to happen.

Table 1: Data Characteristics

	Abbreviation	**Description**	**N**	**N/T**
Data Set 1	SZBM100	100 (10 × 10) portfolios formed on size and BM	100	0.833
Data Set 2	IND48	48 industry portfolios	48	0.400
Data Set 3	SZBM100 + IND48	Combination of SZBM100 and IND48	148	1.233

Table 1 gives a description about the different portfolios we consider, where N denotes the number of assets in the portfolios, T is the number of return observations, and N/T represents the ratio between them. Note that higher N/T ratios are more prone to estimation errors due to multicollinearity.

4 Methodology

4.1 Portfolio Performances

We want to evaluate the out-of-sample performance of the minimum variance portfolio with the risky assets (stocks) as described in the data section. We do not incorporate risk-free assets in the portfolio, which implies that the portfolio requires full initial investments in risky assets. This imposes the familiar portfolio constraint $1_N'w = 1$, i.e. the weights should sum up to one.

For the out-of-sample portfolio performance evaluation, we use the same standard "rolling-horizon" approach as Goto and Xu (2015). We focus on the GMV portfolio, therefore we only need an estimator of the precision matrix Σ^{-1}. So when we want to construct the GMV portfolio for a particular month t, stock returns from the past 10 years (hence the "estimation window is $T = 120$, because we work with monthly returns) are used to obtain an estimator for Σ^{-1}. This estimation window is chosen according to standard practices in the literature. With this estimator the portfolio weights can be computed, and this weights are used to calculate the portfolio returns in month $t + 1$ out-of-sample (i.e. the holding period is one month). For the next iteration, we drop the earliest returns and add the returns for the next period from the data set to get new weights. This iterative procedure is repeated until we acquire portfolio returns for 330 months ($T = 330$).

For a given estimator $\hat{\Sigma}^{-1}$, the weights for the GMV portfolio can be calculated with the following formula:

$$w_{GMV} = \frac{1}{\mathbf{1}_N' \hat{\Sigma}^{-1} \mathbf{1}_N} \hat{\Sigma}^{-1} \mathbf{1}_N. \tag{12}$$

Now we use $\hat{\Psi}_\rho$ (the proposed estimator by Goto and Xu (2015)), $\hat{\Psi}_\lambda$ (our own proposed estimator) and some alternative estimators for $\hat{\Sigma}^{-1}$. This way we can compare the performance of the GMV portfolios constructed with those different estimators using various performance measurements. All the used estimators are summarized in table 2.

Table 2: Estimators For GMV Portfolio

	Abbreviation	**Portfolio Weights**
Glasso	GMV-$\hat{\Psi}_\rho$	$(\mathbf{1}_N' \hat{\Psi}_\rho \mathbf{1}_N)^{-1} \hat{\Psi}_\rho \mathbf{1}_N$
Elastic	GMV-$\hat{\Psi}_\lambda$	$(\mathbf{1}_N' \hat{\Psi}_\lambda \mathbf{1}_N)^{-1} \hat{\Psi}_\lambda \mathbf{1}_N$
Sample-Based	GMV-$\hat{S}^{-1}$	$(\mathbf{1}_N' \hat{S}^{-1} \mathbf{1}_N)^{-1} \hat{S}^{-1} \mathbf{1}_N$
Equal-Weighted	GMV-EW	$(\mathbf{1}_N' \mathbf{1}_N)^{-1} \mathbf{1}_N$
Jagannathan and Ma	GMV-JM	$(\mathbf{1}_N' \hat{S}^{-1} \mathbf{1}_N)^{-1} \hat{S}^{-1} \mathbf{1}_N^{*}$
Ledoit and Wolf	GMV-LW	$(\mathbf{1}_N' \hat{\Sigma}_{LW}^{-1} \mathbf{1}_N)^{-1} \hat{\Sigma}_{LW}^{-1} \mathbf{1}_N$

Table 2 represents the different methods we use to calculate the precision matrix, where the GMV portfolios all have an abbreviation regarding their method. Additionally, we also present the formula which can be used to obtain optimal portfolio weights for the particular methods.

The first two in the table are the already thoroughly explained estimators where this paper focuses on. The third one is the sample-based GMV portfolio, with the common inverse of the sample-covariance matrix as estimator. The fourth one is the equal-weighted portfolio, where no estimation of the precision matrix is required. The fifth one is proposed by Jagannathan and Ma (2003) and also uses the inverse of the sample-covariance matrix S^{-1}, but also imposes a no-short-sale constraint ($*$). Specifically, in their method they minimize $w' \hat{S} w$ subject to $1_N' w = 1$ and $w_i \geq 0$. The last row in the table is a GMV-portfolio with a shrinkage estimator constructed by Olivier Ledoit and Wolf (2004b). Their shrinkage estimator is a convex combination of two positive definite estimators, so therefore positive definite as well. In the rest of the paper we refer to this portfolios by their abbreviation.

4.2 Regularization Parameters ρ and λ

The proposed estimators $\hat{\Psi}_\rho$ and $\hat{\Psi}_\lambda$ depend respectively on the regularization parameters ρ and λ. We use the log predictive Gaussian likelihood function to get an empirical measure for the performance of different values of ρ and λ. To avoid a look-ahead bias, the first 10 year period

(120 months) has been used as in-sample period, starting from July 1963 until July 1973. The log predictive Gaussian likelihood function with a certain precision matrix estimator $\hat{\Sigma}^{-1}$ can be expressed as follows:

$$L(\hat{\Sigma}^{-1}) = \frac{1}{T_f}\sum_{t=1}^{T_f}(ln(det(\hat{\Sigma}_{t-1}^{-1})) - \widetilde{R}_t'\hat{\Sigma}_{t-1}^{-1}\widetilde{R}_t), \tag{13}$$

where T_f is the number of out-of-sample testing periods, which is 120 in this case. $\widetilde{R}_t = R_t - \frac{1}{T_f}$ denotes the demeaned return vector, computed by taking the time series mean of the in-sample training period and subtract it from the out-of-sample return in the testing period, for all individual assets.

To obtain the optimal values of ρ and λ, we start at July 1963 and use the in-sample period consisting of $T = 120$ observations to compute the necessary compartments of the formula (4). This is done in an iterative way for 120 periods, whereafter we arrive at an average predictive likelihood score for different values of respectively ρ and λ. The values of ρ and λ ranged from 0.4 to 3.0, and we used a grid search procedure with increments of 0.1 to arrive at values of ρ and λ that maximize the predictive likelihood. Table 3 shows the founded optimal regularization parameter values with the previously described method, together with the sparsity values for $\hat{\Psi}_\rho$ and $\hat{\Psi}_\lambda$.

Table 3: Regularization Parameters And Sparsity Values

	Abbreviation	$\boldsymbol{\rho}$	$\boldsymbol{\lambda}$	**Sparsity** $\hat{\Psi}_\rho$	**Sparsity** $\hat{\Psi}_\lambda$
Data Set 1	SZBM100	1.1	2.1	75.2%	74.2%
Data Set 2	IND48	1.3	1.9	55.7%	48.8%
Data Set 3	SZBM100 + IND48	1.3	2.7	79.7%	78.2%

Table 3 shows the found optimal regularization parameters for our proposed mean variance optimizers $\hat{\Psi}_\rho$ and $\hat{\Psi}_\lambda$. It also presents the sparsity values, which is the average percentage of the zero values of the off-diagonal elements.

4.3 Performance Measurements

4.3.1 Out-of-Sample Portfolio Risk

The main interest is whether the proposed estimators $\hat{\Psi}_\rho$ (by Goto and Xu (2015)) and $\hat{\Psi}_\lambda$ (by us) are able to reduce the out-of-sample portfolio risk. With the described methods earlier in this section, we acquire GMV weights and out-of-sample portfolio returns for the six used estimators (as shown in Table 2) on the three data sets (described in the data section). With this out-of-sample portfolio returns we can calculate out-of-sample portfolio variances. From this variances we calculate the out-of-sample portfolio standard deviations, denoted by $\sigma_{\hat{\Psi}_\rho}$, $\sigma_{\hat{\Psi}_\lambda}$, $\sigma_{\hat{S}^{-1}}$, σ_{EW}, σ_{JM}, and σ_{LW}.

Now we can compare all the out-of-sample portfolio risks (measured by standard deviations) with each other. The null hypothesis we test is the one of no difference in out-of-sampling portfolio risk. To this end, we compute bootstrap two-sided confidence intervals for $\sigma_{\hat{S}^{-1}} - \sigma_{\hat{\Psi}_i}$, $\sigma_{EW} - \sigma_{\hat{\Psi}_i}$, $\sigma_{JM} - \sigma_{\hat{\Psi}_i}$, $\sigma_{LW} - \sigma_{\hat{\Psi}_i}$, where $i \in \{\rho, \lambda\}$. So each of the proposed estimators $\hat{\Psi}_\rho$ and $\hat{\Psi}_\lambda$ are compared with the four alternative estimators. Lastly, we also compare the two proposed estimators with themselves, i.e. $\sigma_{\hat{\Psi}_\rho} - \sigma_{\hat{\Psi}_\lambda}$. We use a nominal level $1-\alpha$, i.e. when this interval does not contain a 0, the null hypothesis is rejected at the nominal level α. For bootstrapping, we apply the "stationary bootstrap" (i.e., the block resampling with block lengths having a geometric distribution) approach from Politis and Romano (1994), because portfolio standard deviations display serial dependence. The mean block size for each data set is chosen using the optimal block size proposed by Politis and White (2004).

4.3.2 Out-of-Sample Sharpe Ratio

The Sharpe ratio can be obtained by dividing the mean excess return by the standard deviation of a portfolio. It is one of the most widely used performance measures in finance, even tough mean returns are susceptible to estimation errors. If the mean returns stay the same, the Sharpe ratio can improve when the portfolio risk has been reduced. Because the main purpose is to look whether out-of-sample risk reduction can be achieved, we follow the line of Goto and Xu (2015) and calculate out-of-sample Sharpe ratios for all the portfolios we construct.

We also compare the calculated Sharpe ratios in the same way as we compared the standard deviations. However, it should be noted that due to large estimation errors in mean returns, it is difficult to gather reliable differences in the Sharpe ratios. The generally used test proposed by J Dave Jobson and B. M. Korkie (1981) is also not suitable through the presence of fat tails, serial correlation, and volatility clustering. Hence we use the studentized circular block bootstrap method of Oliver Ledoit and Wolf (2008), also with the optimal block size according to the approach of Politis and White (2004). The null hypothesis here is no difference in Sharpe ratios.

4.3.3 Optimized Portfolio Weights

The sample covariance matrix is susceptible to estimation errors, which could cause unstable and extreme valued portfolio weights. This implies that a more reliable estimation method for the sample covariance matrix can possibly help to avoid this problem. We already discussed some methods to estimate the sample covariance matrix in this section. For example, Olivier Ledoit and Wolf (2003), (2004a), (2004b) suggest a shrinkage method, Jagannathan and Ma (2003) propose a method with explicit non-negativity constraints on the portfolio weights themselves, Goto and Xu (2015) take an approach with direct shrinking of the precision matrix. Our own proposed elastic net regression method also shrinks the precision matrix. Therefore, it is worthwhile to compare the optimized portfolio weights and their behavior for the different constructed portfolios.

For the different constructed portfolios with the three data sets, we tabulate the minimum, 1st, 5th, 95th, and 99th percentiles, as well as the maximum of the portfolio weights. We also calculate the Herfindahl index of optimized portfolio weights, denoted by $(1/T)\sum_{t=1}^{T}(\sum_{i=1}^{N}\hat{w}_{i,t}^{2})$, where $\hat{w}_{i,t}^{2}$ is the optimized weight on asset i at period t. When all assets receive equal weights, the Herfindahl index takes the lowest value. This implies that the index takes larger values when the weights are more variable across each other.

We can also look at the monthly variability of the optimized weights over time from each method by looking at the monthly turnover. The turnover can be interpreted as the average fraction of wealth traded in each rebalancing period. We use the formula from DeMiguel, Garlappi, Nogales, et al. (2009) to calculate it, denoted by

$$Turnover = \frac{1}{T-\tau-1}\sum_{t=\tau}^{T-1}\sum_{j=1}^{N}(|w_{j,t+1}^{i} - w_{j,t^{+}}^{i}|). \tag{14}$$

In the definition of turnover, $w_{j,t}^{i}$ denotes the portfolio weight taken in asset j at time t, with used portfolio construction method i. The shown formula consists of $w_{j,t^{+}}^{i}$, the portfolio weight before rebalancing but at $t+1$, and $w_{j,t+1}^{i}$, the portfolio weight at $t+1$ but after rebalancing (i.e. the desired portfolio weight). The interpretation of the turnover, as can be seen by the formula, is that you look at the sum of the absolute differences of the weights before and after rebalancing for every asset j, over $T-\tau-1$ trading dates, normalized by the total number of trading dates.

4.3.4 Economic Gains from Improved Portfolio Optimization

The aim of the proposed methods is to accomplish portfolio risk reduction beyond the EW $(1/N)$ diversification rule in mean variance portfolio optimization. To achieve this we have to rebalance the weights to their optimal weights for every month in the testing period. However, to assess whether the portfolio risk reduction is economically significant, we also have to take into account the transaction costs of this monthly trades. Such costs can grow rapidly with certain methods, like with the hedge-trade-induced turnovers with the method of Goto and Xu (2015).

The measurement we use to evaluate the economic significance is the annualized certainty equivalent excess return (CER) of each portfolio, after adjustment with the transaction costs (T_COST) through subtraction. Specifically, the $T\_COST\_ADJUSTED\_CER$ can be denoted as

$$T\_COST\_ADJUSTED\_CER\_q = \hat{\mu}_q - \frac{\gamma}{2}\hat{\sigma}_q^2 - T\_COST_q, \tag{15}$$

where $\hat{\mu}_q$ and $\hat{\sigma}_q^2$ are the time series of respectively the annualized mean and variance of out-of-sample excess returns for portfolio q. The risk aversion coefficient γ is set at 5, as also done by Brandt (2010). The T_COST_q for portfolio q denotes the annualized turnover costs for that portfolio, measured by the annualized turnover (which we also calculated) multiplied by proportional transaction costs of 50 basis points (bps) per trade. This is according to standard practice in the recent literature. The $T\_COST\_ADJUSTED\_CER\_q$ can be interpreted as an asset that guarantees the risk-free rate plus some (positive) return value, which makes the investor indifferent between holding that asset or a certain risky portfolio q (after accounting for transaction costs). This implies that higher values of the $T\_COST\_ADJUSTED\_CER\_q$ indicate a risk-return characteristic which is more desirable by the investor.

5 Results

5.1 Out-of-Sample Portfolio Risk Minimization

In table 4 we present the out-of-sample variances of the six portfolios for the three data sets we used, denoted in percentages. It can already be seen that in the GMV-$\hat{S}^{-1}$ portfolio the variances increase substantially when the number of assets N approach the number of observations T in magnitude. Furthermore, the GMV-$\hat{S}^{-1}$ portfolio cannot even be constructed when $T < N$, because $\hat{S}$ is singular and non-invertible then, i.e. $\hat{S}^{-1}$ does not exist. For the $GMV - EW$, $GMV - JM$ and $GMV - LW$ portfolios, we see that the variances vary in a more moderate way. When we look at the variances of our two proposed mean variance optimizer portfolios GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$, we see that they generate lower out-of-sample portfolio variances than the alternative portfolios for all data sets.

Table 4: Return Variances (%)

	Abbreviation	$\sigma^2_{\hat{\Psi}_\rho}$	$\sigma^2_{\hat{\Psi}_\lambda}$	$\sigma^2_{\hat{S}^{-1}}$	σ^2_{EW}	σ^2_{JM}	σ^2_{LW}
Data Set 1	SZBM100	13.88	13.84	58.81	25.80	18.37	18.87
Data Set 2	IND48	12.32	12.43	17.31	22.64	13.26	12.94
Data Set 3	SZBM100 + IND48	10.78	10.65	T<N	24.23	12.70	12.04

Table 4 lists the computed monthly return variances (in %), of all our used GMV portfolios, for all the data sets. This has been done over a out-of-sample period of 330 months.

We also used a bootstrap method (as explained in the methodology section) to compare the out-of-sample portfolio risks (measured by the standard deviation) for GMV-$\hat{\Psi}_\rho$ as well as GMV-$\hat{\Psi}_\lambda$ and the alternative portfolios. The results of this comparisons are shown in table 5 and 6 for GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$, respectively. Note that ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. It can be immediately seen that the two tables have quite comparable results. We also see that all the differences are positive and significantly different at the 1% level, which implies that both the GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ generate lower portfolio risks than their alternatives. What also can be noted is that the GMV portfolios where the covariance matrices are obtained via regularization/shrinkage methods (GMV-$\hat{\Psi}_\rho$, GMV-$\hat{\Psi}_\lambda$, GMV-LW) all outperformed the naïve diversification rule (GMV-EW).

Table 5: $\boldsymbol{\sigma_{ALT} - \sigma_{\hat{\Psi}_\rho}}$ (%)

	Abbreviation	$\sigma_{\hat{S}_{-1}}$	σ_{EW}	σ_{JM}	σ_{LW}
Data Set 1	SZBM100	3.95***	1.36***	0.57***	0.62***
Data Set 2	IND48	0.65***	1.24***	0.14***	0.09***
Data Set 3	SZBM100 + IND48	T<N	1.63***	0.29***	0.19***

Table 5 shows the mean differences between the out-of-sample standard deviations (in %) of the alternative used portfolios (except GMV-$\hat{\Psi}_\lambda$), and the GMV-$\hat{\Psi}_\rho$ portfolio. We used the stationary bootstrap method of Politis and Romano (1994), with optimal expected block size computed with a method proposed by Politis and White (2004). The null hypothesis tested is the one of no difference in the two-sided bootstrap intervals, where ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table 6: $\boldsymbol{\sigma_{ALT} - \sigma_{\hat{\Psi}_\lambda}}$ (%)

	Abbreviation	$\sigma_{\hat{S}_{-1}}$	σ_{EW}	σ_{JM}	σ_{LW}
Data Set 1	SZBM100	3.95***	1.36***	0.57***	0.62***
Data Set 2	IND48	0.64***	1.23***	0.12***	0.08***
Data Set 3	SZBM100 + IND48	T<N	1.65***	0.31***	0.21***

Table 6 shows the same as Table 5, but now for GMV-$\hat{\Psi}_\lambda$.

5.2 Out-Of-Sample Sharpe Ratio

In table 6 we present the monthly Sharpe ratios for all the GMV portfolios. We see that GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ generate the highest Sharpe ratios compared to the alternative portfolios for data set 1 and data set 3. For data set 2, GMV-EW and GMV-JM attain higher Sharpe ratios than GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$, but they are still considerably close. The Sharpe ratios are calculated for the testing period between July 1983 and Dec. 2010 and range between 0.147 and 0.211 for GMV-$\hat{\Psi}_\rho$ and between 0.140 and 0.229 for GMV-$\hat{\Psi}_\lambda$. To get a more intuitive feeling for this Sharpe ratios, note that the value-weighted U.S. market portfolio had a Sharpe ratio of 0.113 during the same period.

Table 7: Monthly Sharpe Ratios

	Abbreviation	$\hat{\boldsymbol{\Psi}}_{\boldsymbol{\rho}}$	$\hat{\boldsymbol{\Psi}}_{\boldsymbol{\lambda}}$	$\hat{\boldsymbol{S}}^{-1}$	***EW***	***JM***	***LW***
Data Set 1	SZBM100	0.211	0.211	0.127	0.104	0.113	0.150
Data Set 2	IND48	0.147	0.140	0.097	0.174	0.148	0.133
Data Set 3	SZBM100 + IND48	0.210	0.229	T<N	0.127	0.171	0.193

Table 7 presents the monthly out-of-sample sharpe ratios of all the used portfolios, for all the data sets.

The results of the difference of Sharpe ratios between our proposed mean variance optimizers and the alternative portfolios, computed with the studentized circular block bootstrap approach from Oliver Ledoit and Wolf (2008), are tabulated in table 5 and table 6 for respectively GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$. Note that ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. We find that for data set 1, GMV-$\hat{\Psi}_\rho$ as well as GMV-$\hat{\Psi}_\lambda$ significantly (at 5% level) higher Sharpe ratios than GMV-EW and GMV-JM. For data set 3, the Sharpe ratios of GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ are also significantly higher than GMV-EW at respectively the 10% and 5% level. Data set 2 does not seem to give significantly different results between Sharpe ratios of our used portfolios. The difference between GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ themselves is not statistically significant (not presented in the table). Note that finding reliable differences is hard due to large estimation errors when calculating the mean returns needed for computation of the Sharpe ratios. This implies that we cannot make strong conclusions about this particular results.

Table 8: Difference SRs GMV-$\hat{\Psi}_\rho$ And Alternative Portfolios

	Abbreviation	$\hat{\boldsymbol{S}}^{-1}$	***EW***	***JM***	***LW***
Data Set 1	SZBM100	0.083	0.107**	0.098**	0.063
Data Set 2	IND48	0.049	-0.028	-0.002	0.014
Data Set 3	SZBM100 + IND48	T<N	0.083*	0.039	0.017

Table 8 lists the differences between monthly Sharpe ratios between GMV-$\hat{\Psi}_\rho$ and the alternative used portfolios (except GMV-$\hat{\Psi}_\lambda$). We used the studentized circular block bootstrap method of Oliver Ledoit and Wolf (2008) with an optimal block size. The null hypothesis tested is the one of no difference in the two-sided bootstrap intervals, where ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table 9: Difference SRs GMV-$\hat{\Psi}_\lambda$ And Alternative Portfolios

	Abbreviation	$\hat{\boldsymbol{S}}^{-1}$	***EW***	***JM***	***LW***
Data Set 1	SZBM100	0.084	0.108**	0.098**	0.064
Data Set 2	IND48	0.042	-0.035	-0.009	0.007
Data Set 3	SZBM100 + IND48	T<N	0.101**	0.058	0.036

Table 9 shows the same as Table 8, but now for GMV-$\hat{\Psi}_\lambda$.

5.3 Behavior of Optimal Portfolio Weights

In table 10 we tabulate the 0% (minimum), 1%, 5%, 95% and 100% (maximum) percentiles of the optimized weights of all the used portfolios for the three data sets. The trivial GMV-EW portfolio weights are excluded from this table. GMV-$\hat{S}^{-1}$ takes quite extreme weights for data sets 1 and 2, with a range from -2.246 to 1.826 for data set 1, and a range from -0.707 and 0.875 for data set 2. Especially in data set 1, GMV-$\hat{S}^{-1}$ seems to have far more extreme weights compared with the other portfolios. It can also be seen that the GMV-JM portfolio with the no-short-sale constraint tends to produce portfolios which are highly concentrated, because the 95% value is very low in all the three data sets. The portfolios which use hedge trades (GMV-$\hat{\Psi}_\rho$, GMV-$\hat{\Psi}_\lambda$, and GMV-LW) use both long- and short positions and do not seem to employ weights which are that extreme.

Table 10: Distribution of Portfolio Weights

	Abbreviation	**Portfolio**	**Min**	**1 %**	**5 %**	**95 %**	**99 %**	**Max**
Data Set 1	SZBM100	GMV-$\hat{\Psi}_\rho$	-0.243	-0.146	-0.105	0.145	0.208	0.305
		GMV-$\hat{\Psi}_\lambda$	-0.246	-0.147	-0.106	0.146	0.209	0.306
		GMV-$\hat{S}^{-1}$	-2.246	-0.720	-0.494	0.526	0.820	1.826
		GMV-JM	0	0	0	0.080	0.231	0.461
		GMV-LW	-0.198	-0.132	-0.092	0.143	0.205	0.353
Data Set 2	IND48	GMV-$\hat{\Psi}_\rho$	-0.235	-0.151	-0.104	0.159	0.379	0.518
		GMV-$\hat{\Psi}_\lambda$	-0.245	-0.166	-0.110	0.164	0.401	0.543
		GMV-$\hat{S}^{-1}$	-0.707	-0.359	-0.200	0.246	0.524	0.875
		GMV-JM	0	0	0	0.109	0.523	0.717
		GMV-LW	-0.256	-0.148	-0.103	0.174	0.490	0.664
Data Set 3	SZBM100 + IND48	GMV-$\hat{\Psi}_\rho$	-0.152	-0.092	-0.056	0.075	0.114	0.244
		GMV-$\hat{\Psi}_\lambda$	-0.178	-0.108	-0.068	0.091	0.142	0.284
		GMV-$\hat{S}^{-1}$			T <N			
		GMV-JM	0	0	0	0.023	0.149	0.700
		GMV-LW	-0.186	-0.110	-0.077	0.111	0.185	0.423

Table 10 lists the distribution of the monthly weights of our portfolios (except GMV-EW). We present the 0% (minimum), 1%, 5%, 95% and 100% (maximum) percentiles of the optimized weights, for all the data sets.

We present the calculated Herfindahl indices in table 11. The index takes larger values when the weights are more variable across each other. So as already expected, the Herfindahl index for the GMV-$\hat{S}^{-1}$ portfolio takes large values, which we denote with VLN (very large number). Furthermore, the Herfindahl indices for the hedge regression portfolios (GMV-$\hat{\Psi}_\rho$, GMV-$\hat{\Psi}_\lambda$, and GMV-LW) for data set 1 and 2 seem quite comparable. The GMV-JM portfolio has the lowest Herfindahl indices for all the data sets, which could be due to the fact that they entail highly concentrated portfolios with relatively less assets hold compared to the other portfolios.

Table 11: Herfindahl Index

	Abbrev.	Portfolio	Herfindahl index (%^{2})		Abbrev.	Portfolio	Herfindahl index (%^{2})
Data 1	SZBM100	GMV-$\hat{\Psi}_\rho$	5,923	**Data 2**	IND48	GMV-$\hat{\Psi}_\rho$	4,122
		GMV-$\hat{\Psi}_\lambda$	5,961			GMV-$\hat{\Psi}_\lambda$	4,440
		GMV-$\hat{S}^{-1}$	VLN			GMV-$\hat{S}^{-1}$	VLN
		GMV-JM	1,810			GMV-JM	3,058
		GMV-LW	5,439			GMV-LW	5,223
Data 3	SZBM100 + IND48	GMV-$\hat{\Psi}_\rho$	2,638				
		GMV-$\hat{\Psi}_\lambda$	3,694				
		GMV-$\hat{S}^{-1}$	T<N				
		GMV-JM	2,883				
		GMV-LW	5,099				

Table 11 lists the Herfindahl Index of our portfolios (except GMV-EW). The more variable the weights are across each other, the higher the index value.

The monthly portfolio turnovers are presented in table 12. It is clearly that the GMV-EW portfolio has the lowest turnovers for all the data sets, because with this naïve diversification rule only small rebalancing trades have to be made every month. The GMV-JM portfolio also has low turnovers, likely due to the no-short-sale constraint. It can also be seen that the portfolios who employ hedge trades (GMV-$\hat{\Psi}_\rho$, GMV-$\hat{\Psi}_\lambda$, GMV-LW) contain a higher turnover. However, the GMV-$\hat{S}$ portfolio attains the highest turnover for all data sets, with an average monthly turnover of 7.005 for data set 1.

Table 12: Portfolio Turnover (Monthly)

	Abbreviation	$\hat{\Psi}_\rho$	$\hat{\Psi}_\lambda$	$\hat{S}^{-1}$	EW	JM	LW
Data Set 1	SZBM100	0.561	0.564	7.005	0.024	0.133	0.588
Data Set 2	IND48	0.297	0.326	0.775	0.033	0.077	0.329
Data Set 3	SZBM100 + IND48	0.465	0.601	T<N	0.027	0.092	0.769

Table 12 reports the monthly turnover of the used portfolios, following the calculation method of DeMiguel, Garlappi, Nogales, et al. (2009). $T < N$ indicates that the sample covariance matrix is invertible, thus the portfollio cannot be constructed.

5.4 Economic Gains from Improved Portfolio Optimization

In table 9 we report the T_COST-ADJUSTED_CERs of the six portfolios. Our proposed GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ portfolios clearly accomplish better economic gains than the GMV-$\hat{S}$ portfolio. Moreover, although GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ have much larger transaction costs due to the hedge trades which need to be made, they outperform GMV-EW and GMV-JM portfolios (which have less transaction costs) in many data sets due to the gains in risk reduction. This implies that in many situations for the investor with a risk-aversion coefficient of $\gamma = 5$, the reduced portfolio risk outweighs the transaction costs. If we look at the hedge regression portfolios and compare them with each other, we see that GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ both have more favorable economic gains than the GMV-LW portfolio of Olivier Ledoit and Wolf (2004b), for all data sets. The economic gains of GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ do not vary that much from each other.

Table 13: T_COST-ADJUSTED_CERs (annual %)

	Abbreviation	$\hat{\boldsymbol{\Psi}}_{\boldsymbol{\rho}}$	$\hat{\boldsymbol{\Psi}}_{\boldsymbol{\lambda}}$	$\hat{\boldsymbol{S}}^{\mathbf{-1}}$	EW	JM	LW
Data Set 1	SZBM100	2.30	2.32	-47.29	-1.37	-0.34	-1.21
Data Set 2	IND48	0.88	0.39	-4.87	3.44	2.24	0.02
Data Set 3	SZBM100 + IND48	2.58	2.52	T<N	0.34	3.19	0.12

Table 13 lists the T_COST-ADJUSTED_CERs of the used portfolios, shown in annual percentage points. Transaction costs are calculated as 50 bps times monthly turnover times 12 (to annualize).

5.5 Differences with Goto and Xu (2015)

Many of our results are quite comparable with the paper of Goto and Xu (2015), but some of them are different. This is likely due to the fact that the data sets which have been used were modified last year. It could also be because the methods we used to obtain the results differed in some aspects compared to the methods Goto and Xu (2015) used. Note that we do not intend to claim that our methods or results are superior to the ones of Goto and Xu (2015). On the contrary, our aim was to replicate their methods and come up with the same results, so we acknowledge our possible shortcomings.

6 Conclusion

The aim of this paper is to use the insight that the precision matrix contains optimal hedging relations among stocks, which allows us to reduce estimation errors by shrinkage methods. The proposed mean variance optimizers, Ψ_ρ and Ψ_λ, both achieve sparsity, i.e. a significant part of their off-diagonal elements are set to 0.

We have shown that the GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ portfolios both accomplish significant and robust out-of-sample risk reduction compared to the standard GMV-$\hat{S}^{-1}$ portfolio, especially when the sample covariance matrix is singular. We also see that the out-of-sample performance of the GMV-$\hat{\Psi}_\rho$ and GMV-$\hat{\Psi}_\lambda$ portfolios compare favorably to the portfolios GMV-EW, GMV-JM and GMV-LW in many situations. Furthermore, while the turnover costs are quite substantial for our proposed portfolios, the gains in risk reduction outweigh this costs in many situations.

The results are in line with the theoretical motivation of why we wanted to implement this methods in the first place. By mitigating the estimation errors in the hedging relations of the precision matrix, the proposed mean variance optimizers can significantly reduce the out-of-sample portfolio risk.

References

[1] Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data". In: *The Journal of Machine Learning Research* 9 (2008), pp. 485–516.

[2] Michael W Brandt. "Portfolio choice problems". In: *Handbook of financial econometrics: Tools and techniques*. Elsevier, 2010, pp. 269–336.

[3] Laurent Callot, Mehmet Caner, A Özlem Önder, and Esra Ulaşan. "A nodewise regression approach to estimating large portfolios". In: *Journal of Business & Economic Statistics* (2019), pp. 1–12.

[4] Victor DeMiguel, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal. "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms". In: *Management science* 55.5 (2009), pp. 798–812.

[5] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *The review of Financial studies* 22.5 (2009), pp. 1915–1953.

[6] Jerome Friedman, Trevor Hastie, Holger Höfling, Robert Tibshirani, et al. "Pathwise coordinate optimization". In: *Annals of applied statistics* 1.2 (2007), pp. 302–332.

[7] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. "Sparse inverse covariance estimation with the graphical lasso". In: *Biostatistics* 9.3 (2008), pp. 432–441.

[8] Shingo Goto and Yan Xu. "Improving mean variance optimization through sparse hedging restrictions". In: *Journal of Financial and Quantitative Analysis* (2015), pp. 1415–1441.

[9] Arthur E Hoerl and Robert W Kennard. "Ridge regression: applications to nonorthogonal problems". In: *Technometrics* 12.1 (1970), pp. 69–82.

[10] Ravi Jagannathan and Tongshu Ma. "Risk reduction in large portfolios: Why imposing the wrong constraints helps". In: *The Journal of Finance* 58.4 (2003), pp. 1651–1683.

[11] J Dave Jobson and Bob M Korkie. "Performance hypothesis testing with the Sharpe and Treynor measures". In: *Journal of Finance* (1981), pp. 889–908.

[12] J David Jobson and Robert M Korkie. "Putting Markowitz theory to work". In: *The Journal of Portfolio Management* 7.4 (1981), pp. 70–74.

[13] Raymond Kan and Guofu Zhou. "Optimal portfolio choice with parameter uncertainty". In: *Journal of Financial and Quantitative Analysis* (2007), pp. 621–656.

[14] Solt Kovács, Tobias Ruckstuhl, Helena Obrist, and Peter Bühlmann. "Graphical Elastic Net and Target Matrices: Fast Algorithms and Software for Sparse Precision Matrix Estimation". In: *arXiv preprint arXiv:2101.02148* (2021).

[15] MO Kuismin, JT Kemppainen, and MJ Sillanpää. "Precision matrix estimation with ROPE". In: *Journal of Computational and Graphical Statistics* 26.3 (2017), pp. 682–694.

[16] Oliver Ledoit and Michael Wolf. "Robust performance hypothesis testing with the Sharpe ratio". In: *Journal of Empirical Finance* 15.5 (2008), pp. 850–859.

[17] Olivier Ledoit and Michael Wolf. "A well-conditioned estimator for large-dimensional covariance matrices". In: *Journal of multivariate analysis* 88.2 (2004), pp. 365–411.

[18] Olivier Ledoit and Michael Wolf. "Honey, I shrunk the sample covariance matrix". In: *The Journal of Portfolio Management* 30.4 (2004), pp. 110–119.

[19] Olivier Ledoit and Michael Wolf. "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection". In: *Journal of empirical finance* 10.5 (2003), pp. 603–621.

[20] Harry Markowitz. "Portfolio Selection". English. In: *The Journal of Finance* 7.1 (Mar. 1952), pp. 77–91. DOI: 10.2307/2975974. URL: https://www.jstor.org/stable/2975974.

[21] Nicolai Meinshausen, Peter Bühlmann, et al. "High-dimensional graphs and variable selection with the lasso". In: *Annals of statistics* 34.3 (2006), pp. 1436–1462.

[22] Dimitris N Politis and Joseph P Romano. "The stationary bootstrap". In: *Journal of the American Statistical association* 89.428 (1994), pp. 1303–1313.

[23] Dimitris N Politis and Halbert White. "Automatic block-length selection for the dependent bootstrap". In: *Econometric reviews* 23.1 (2004), pp. 53–70.

[24] Adam J Rothman, Peter J Bickel, Elizaveta Levina, Ji Zhu, et al. "Sparse permutation invariant covariance estimation". In: *Electronic Journal of Statistics* 2 (2008), pp. 494–515.

[25] Guy VG Stevens. "On the inverse of the covariance matrix in portfolio analysis". In: *The Journal of Finance* 53.5 (1998), pp. 1821–1827.

[26] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996), pp. 267–288.

[27] Ryan J Tibshirani et al. “The lasso problem and uniqueness”. In: *Electronic Journal of statistics* 7 (2013), pp. 1456–1490.

[28] Wessel N Van Wieringen and Carel FW Peeters. “Ridge estimation of inverse covariance matrices from high-dimensional data”. In: *Computational Statistics & Data Analysis* 103 (2016), pp. 284–303.

[29] Ming Yuan and Yi Lin. “Model selection and estimation in the Gaussian graphical model”. In: *Biometrika* 94.1 (2007), pp. 19–35.

[30] Hui Zou and Trevor Hastie. “Regularization and variable selection via the elastic net”. In: *Journal of the royal statistical society: series B (statistical methodology)* 67.2 (2005), pp. 301–320.