Erasmus University Rotterdam - Erasmus School of Economics

Bachelor Thesis: Econometrie & Operationele Research

Sparse Precision Matrix Estimation with Elastic Net Penalty

Abstract

This research extends on the findings of Goto and Xu (2015) who proposed an improved graphical lasso (glasso) estimator of the precision matrix. This estimator is compared to three estimation methods using elastic net regularisation, instead of lasso regularisation. The graphical elastic net estimator by Kovács et al. (2021) is very similar to the glasso estimator, but for one of the data sets a significant improvement of the out-of-sample Sharpe ratio is found. The node-wise regression estimator is outperformed by both the gelnet and glasso estimator. The 2 stage gelnet estimator by Bernardini et al. (2021) reduced to the gelnet estimator, which makes the 2 stages redundant.

Student: Bhodi Stuivenberg

Student id: 476037

Supervisor: S.O.R. Lonn

Second Assessor: P.A. Opschoor

Date final version: 04/07/2021

Contents

1 Introduction

The estimation of the (inverse) covariance matrix Σ in mean-variance portfolio optimisation is one of the most important aspects of creating minimum variance portfolios. Mean-variance portfolio optimisation was originally introduced by Markowitz (1952), who shows that a portfolio of risky assets with minimum variance can be created by including multiple stocks. In the pursuit of this mean-variance portfolio the key aspect is the estimation of Σ^{-1}, because it is used to determine the optimal portfolio weights. By allocating lower weights to riskier stocks (stocks with higher variance) and larger weights to less risky stocks, the variance of the portfolio returns can be reduced.

In practice the estimation of this precision matrix has proven to be difficult. Demiguel et al. (2009) show that 14 optimised models in their research do not consistently outperform the naive equal weighted portfolio, due to large estimation errors. They conclude that the gain of optimising portfolios is cancelled out by the large estimation errors that exist when estimating the parameters, such as the precision matrix. More recent research by Hwang et al. (2017) shows that the naive equal weighted portfolio outperforms the mean-variance portfolio. They also found that increasing the number of stocks leads to larger estimation errors, because more parameters need to be estimated.

Two factors that lead to the large estimation errors are: high correlations between the stock returns and the limited amount of observations (T) compared to the amount of stocks (N). It is well known that stock returns are highly correlated, which causes multi-collinearity. Multi-collinearity leads to large estimation errors of the parameters, which are in this case the elements of the (inverse) covariance matrix. These elements are often estimated by ordinary least squares (OLS) regressions when $N < T$, because regression can be used to find the optimal hedging relations between stocks (Stevens, 1995). Using OLS will lead to large estimation errors when multi-collinearity is present.

Literature suggests that using different techniques, such as shrinkage methods and factor models applied to precision matrix Σ^{-1}, could lead to improvements of the mean-variance portfolio. Ledoit and Wolf (2004) show that their shrinkage method, which shrinks the precision matrix towards the constant correlation matrix by using a weighted average of the covariance matrix and the constant correlation matrix, can result in higher returns with lower variance compared to using the sample precision matrix. J. Fan et al. (2008) found that using a factor model for estimating the covariance matrix is advantageous compared to using the sample covariance matrix.

This research uses the estimator proposed by Goto and Xu (2015), who apply the graphical lasso (glasso) method by Friedman et al. (2008) to the estimation of the precision matrix. The graphical

lasso applies lasso regularisation to the log likelihood of the precision matrix. Lasso regularisation, introduced by Tibshirani (1996), penalises the objective function in such a way that it leads to a more sparse precision matrix, which means that it promotes parameter estimates equal to zero.

The estimation by Goto and Xu (2015) aims to overcome the problem of multi-collinearity by estimating a sparse precision matrix and reducing the magnitude of extreme hedging positions by using the lasso penalty. Their method differs from most regularisation methods, because they regularise the precision matrix directly, instead of the covariance matrix or weight vector.

Zou and Hastie (2005) show that the lasso regularisation can be generalised to the elastic net for linear regression problems. This regularisation technique can be implemented in the same way as lasso regularisation in the method of Goto and Xu (2015). Elastic net is a linear combination of lasso (l_1) regularisation and ridge (l_2) regularisation, which are different types of penalisation functions. Tibshirani (1996) found that empirically l_2 regularisation improves predictive performance compared to l_1 regularisation when the data has high correlations between the predictors and $T > N$. A downside of l_2 regularisation is that it does not promote sparsity, because it does not set elements that are smaller than the regularisation parameter to zero, which the lasso does. Higher sparsity should lead to smaller estimation errors, because less parameters need to be estimated. These findings suggests that the elastic net may outperform l_1 and l_2 regularisation, because it can use the benefits of both techniques. This leads to the main research question: *"Can the performance of the improved mean-variance estimator be improved by using an elastic net penalty?"* where the improved mean-variance estimator is the estimator proposed by Goto and Xu (2015).

Kovács et al. (2021) derived a method to adjust the graphical lasso by Friedman et al. (2008) to include elastic net penalty. This combines the work of Goto and Xu (2015) with the literature on the elastic net. To answer the main research question the following question needs to be answered: *"Can the graphical elastic net improve the performance of the improved mean-variance estimator?"*.

Due to the limited amount of work on graphical elastic net, there might be intricacies that are not found yet. For example, Banerjee et al. (2008) show that the graphical lasso produces estimates that are invertable, but in this research cases where found where the determinant of the precision matrix was 0, which implies non-invertability. Therefore a node-wise regression approach, similar to that of Callot et al. (2021) is used. This method is more easily adaptable to include the elastic net penalty. This leads to the second research question: *"Can node-wise regression with elastic net penalty improve performance of the improved mean-variance estimator?"*.

Bernardini et al. (2021) propose a combination of the graphical elastic net and node-wise re-

gression and show that it improves the identification of relations between variables. They call this combination two stage graphical elastic net (2S Gelnet). This leads to the final research question: *"Can 2 stage graphical elastic net improve performance of the improved mean-variance estimator?"*.

This research adds to the current literature, because it uses the methodology of Goto and Xu (2015) in that it directly estimates the precision matrix via a graphical method and node-wise regression. The graphical method is changed to use a different penalisation function, which could lead to an increase in portfolio performance. This increase might be relevant to portfolio managers, because it might create better performing mean-variance portfolios. Academically it is also interesting to see if the graphical elastic net could be a new solid estimation method.

The remainder of this paper is structured as follows: Section 2 discusses relevant literature on the topic of precision matrix estimation. Section 3 describes the used data sets. Section 4 explains the different methods used to estimate precision matrices. Section 5 reports the results found by using these methods and Section 6 discusses the concluding remarks on the results.

2 Literature

In the context of regularised covariance matrix estimates there are a lot of shrinkage estimators. Jagannathan and Ma (2003) found that the performance of an estimator which could not include short positions equals that of factor models and other shrinkage methods. They argue that using "wrong constraints" like the no short sales constraint, which leads to misspecification bias, is justified, due to smaller estimation errors. J. Fan et al. (2008) use the Fama-French 3 factor model by Fama and French (1993) to show that using factor models leads to performance improvements in portfolio optimisation. Nard et al. (2018) recently applied new methods based on large-dimensional asymptotic theory. They found that their factor model performed better than the Fama-French 5 factor model, which is known to perform well compared to equal weighted portfolios. Ledoit and Wolf (2004) use a linear combination of a factor structured estimator and the sample covariance to achieve shrinkage in the covariance matrix.

The main interest in this research is the estimator by Goto and Xu (2015) who use lasso regularisation to create a more sparse precision matrix. They use a lasso penalty to apply shrinkage and promote sparsity in the precision matrix. The goal of this paper is to investigate whether changing the lasso penalty to an elastic net penalty improves portfolio performance. Zou and Hastie (2005) mention that l_1 regularisation lacks the ability to set "correct" parameter estimates when 2 or more

variables have very high correlation. Lasso regularisation will choose one of the variables that are highly correlated and set the other to zero, which should not be the case. They also note that elastic net does not have this problem. Economically adding an elastic net penalty also makes sense, because it may lead to a more stable portfolio compared to lasso regularised portfolios, due to the larger penalisation of extreme portfolio weights. Smaller portfolio weights are generally preferred, because extreme positions are often found by the mean-variance portfolio due to estimation errors (Michaud, 1989). These estimation errors lead to unstable portfolios, therefore Michaud (1989) note that there should be constraints that prevent these problems.

Literature has proven that using the elastic net penalty in the context of portfolio optimisation can be beneficial. Ho et al. (2015) found that using an elastic net penalty on the mean-variance criterion can increase the Sharpe ratio of a portfolio when the regularisation parameters are optimised right. Their method is different from this research in the way that they penalise the portfolio weights instead of the elements of the precision matrix. Li (2015) found that an estimator based on hedging regressions with elastic net regularisation results in the highest Sharpe ratio on foreign exchange rates and US stocks. Note that using these regressions is different from this research, because the optimal hedging relations are calculated via N separate regressions. In this research the entire precision matrix is estimated at once. Even though the implementations of the elastic net are different in these articles, they suggest that using an elastic net penalty can be beneficial.

3 Data

In order to compare with the results of Goto and Xu (2015) a selection of the same data sets are used. The 3 data sets are available on French's Data Library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The first data set contains monthly returns of 100 portfolios formed on size and book-to-market ratio on the U.S. market. The second data set contains monthly returns of 48 industry portfolios on the U.S. market. The third data set is the combination of the first two data sets. Table 1 reports the N/T ratio and time period for every data set. Reference names are the same as in Goto and Xu (2015) to easily compare results when referring to the data sets. N/T is the number of stocks divided by the number of months in the estimation sample (120 months). Goto and Xu (2015) chose these data sets, because of the high N/T ratio, which causes more problems due to larger estimation errors. Missing values are replaced by the value-weighted market return, available on French's Data Library, following Jagannathan and Ma (2003).

Table 1: Data Description

Reference	Description	N/T	Period
SZBM100	100 portfolios formed on size and book-to-market ratio	0.833	07/1963-12/2010
IND48	48 industry portfolios	0.400	07/1963-12/2010
SZBM100+IND48	Combination of SZBM100 and IND48	1.233	07/1963-12/2010

The period mentioned in Table 1 is split in 2 parts: a training and an out-of-sample period. The training period contains the first 240 observations (July 1963 - June 1983). In this training period the parameters of the estimation methods are optimised, based on 120 estimators (July 1973 - June 1983). These estimators are calculated based on the past 120 months, therefore there are 120 estimators to train the parameters based on 240 months of data. The performance of the different portfolios is compared in the out-of-sample period between July 1983 and December 2010.

4 Methodology

In this section the three main methods and the concept of regression hedging are explained. The three main methods are: graphical lasso, graphical elastic net and node-wise regression. The 2 stage graphical elastic net estimator is an extension of graphical elastic net and node-wise regression, because it uses both methods to estimate the precision matrix.

4.1 Regression hedging

Stevens (1995) found that regression can be used to estimate the precision matrix Σ^{-1}. Regressing stock returns of stock i on the remaining $N-1$ stocks leads to optimal hedging relations, which can be used to estimate the precision matrix. The regression of stock returns follows the notation of Goto and Xu (2015) and is defined as follows:

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

where $r_{i,t}$ is the return of stock i in period t, $\epsilon_{i,t}$ is the error of stock i in period t which is the unhedgeable component of stock i in period t and $\beta_{i|k}$ is the marginal 'effect' of stock k on stock i. Since $\beta_{i|k}$ shows how strong the relation between stock i and k is, Stevens (1995) shows that it can be used for determining hedging relationships when scaled with the variance of $\epsilon_{i,t}$. Define

$v_i = var(\epsilon_{i,t})$ as a measure of the unhedgeable risk of stock i, so the relation to the precision matrix from Stevens (1995) is defined as follows:

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

where $\Sigma_{i,j}^{-1}$ is element (i, j) of Σ^{-1}. This allows for precision matrix estimation via hedge regressions, where every row of Σ^{-1} represents a hedging relation between asset i and the other $N - 1$ assets.

4.2 Goto and Xu's Improved Mean Variance Optimizer

In financial data it is well known that stock returns can be highly correlated, which may lead to multicollinearity when regressing the returns of different stocks on each other. This will lead to large estimation errors in the hedge regression, which will lead to large estimation errors in the precision matrix. Multicollinearity is also caused by the fact that the number of observations can be close to the amount of stocks in practical applications. This form of multicollinearity also leads to unstable estimations with large estimation errors when using hedge regressions.

To overcome this multicollinearity Goto and Xu (2015) implement l_1 regularisation, which is also known as lasso regularisation, to the hedge regression, which means that $\beta_{i|k}$ is estimated by the following equation:

$$\hat{\beta}_{i|k}^{lasso} = \arg\min_{\beta} \left\{ \sum_{t=1}^{T} \left(r_{i,t} - \sum_{k=1,k\neq i}^{N} \beta_{i|k} r_{k,t} \right)^2 + \gamma \sum_{k=1,k\neq i}^{N} |\beta_{i|k}| \right\} \tag{3}$$

where γ is a soft threshold, which means that when $|\hat{\beta}_{i|k}^{OLS}| < \gamma$, then $\hat{\beta}_{i|k}^{lasso} = 0$. Setting these parameters to zero leads to a more sparse hedging portfolio, because if $\beta_{i|k}$ is set to zero it means stock k is not included in the hedging relation of stock i. Using l_1 regularisation also leads to smaller coefficients, because γ is subtracted from all $\beta_{i|k}$.

This l_1 regression would have a clear relation to the OLS estimates if the regressors are orthonormal, but since the regressors are almost never orthonormal another method is needed to calculate $\hat{\beta}_{i|k}^{lasso}$. This problem is solved by Friedman et al. (2007) who use quasi-maximum likelihood (QML) to estimate the entire inverse covariance matrix, instead of each element of the inverse covariance matrix separately. The l_1 norm is implemented in QML by subtracting the same part that was

added to the OLS estimation as in Equation 3. So Σ^{-1} is calculated such that it maximises:

$$\underset{\Sigma^{-1}=[\Sigma_{ij}]}{max} \frac{T}{2}\ln\left(\det\left(\Sigma^{-1}\right)\right) - \frac{T}{2}\text{trace}\left(\hat{S}\Sigma^{-1}\right) - \rho \sum_{i=1,i\neq j}^{N} \sum_{j=1,j\neq i}^{N} |\Sigma_{ij}^{-1}| \tag{4}$$

where $\rho \geq 0$ is the lasso parameter, which penalises the magnitude of the elements of the precision matrix $|\Sigma_{ij}^{-1}|$, by subtracting $\rho \sum_{i=1,i\neq j}^{N} \sum_{j=1,j\neq i}^{N} |\Sigma_{ij}^{-1}|$ from the objective function. It is also a soft threshold, such that if $\Sigma_{ij}^{-1} < \rho$ then Σ_{ij}^{-1} is set to zero. $\hat{S}$ is the sample covariance matrix and det and trace are the matrix determinant and the trace respectively. Note that ρ has the same function as γ in Equation 3. In this research ρ is optimised using the predictive likelihood, following the method of Goto and Xu (2015). More on the optimisation of the penalty parameter is discussed in Section 4.7. The details on the maximisation of Equation 4 can be found in Appendix A.1.

4.3 Graphical elastic net

In the context of regularisation not only l_1 regularisation is a popular method to shrink the number of regressors and the magnitude of the parameters. The l_2 norm, defined as follows:

$$\rho \sum_{i=1,i\neq j}^{N} \sum_{j=1,j\neq i}^{N} (\Sigma_{ij}^{-1})^2 \tag{5}$$

penalises large coefficients more heavily than the l_1 norm for elements larger than 1, since it subtracts the square of the coefficients from the objective function instead of the absolute value. Using l_2 regularisation introduced for linear regression problems by Hoerl and Kennard (1970), which is also called ridge regression, has shown empirically to improve prediction performance. Tibshirani (1996) found that using ridge regression for linear regression problems outperforms the lasso by a substantial amount when there is a high correlation between the independent variables of a model. Financial return data is known to have high correlations between the returns of the different stocks, so using the l_2 norm on this data might improve the predictive performance.

The drawback of using l_2 regularisation instead of l_1 regularisation is that is does not force small parameters to be zero in matrix estimation or linear models. The precision matrix, and thus the hedging portfolios, will therefore likely not be sparse. Sparse hedging portfolios are preferred to minimise estimation errors, by limiting the amount of non-zero elements in the precision matrix.

To overcome the problem of non-sparsity when using the l_2 norm, the elastic net by Zou and Hastie (2005) can be used. They introduced the elastic net penalty for linear regression problems,

which is a linear combination of lasso and ridge regression. The elastic net is defined as follows:

$$\alpha\rho \sum_{i=1,i\neq j}^{N} \sum_{j=1,j\neq i}^{N} |\Sigma_{ij}^{-1}| + (1-\alpha)\rho \sum_{i=1,i\neq j}^{N} \sum_{j=1,j\neq i}^{N} (\Sigma_{ij}^{-1})^2 \tag{6}$$

where $0 \leq \alpha \leq 1$ is a parameter that assigns weights to the l_1 and l_2 norms and ρ determines the magnitude of the l_1 and l_2 penalty. It is of interest whether implementing the elastic net penalty to the mean variance estimator of Goto and Xu (2015) as follows:

$$\begin{aligned} \max_{\Sigma^{-1}=[\Sigma_{ij}]} & \frac{T}{2}\ln\left(\det\left(\Sigma^{-1}\right)\right) - \frac{T}{2}\text{trace}\left(\hat{S}\Sigma^{-1}\right) \\ & - \alpha\rho \sum_{i=1,i\neq j}^{N} \sum_{j=1,j\neq i}^{N} |\Sigma_{ij}^{-1}| - (1-\alpha)\rho \sum_{i=1,i\neq j}^{N} \sum_{j=1,j\neq i}^{N} (\Sigma_{ij}^{-1})^2 \end{aligned} \tag{7}$$

leads to noticeable performance improvements, due to the possible advantages of combining l_1 and l_2 regularisation, which changes the graphical lasso to graphical elastic net.

Little research has been done in the field of graphical elastic net, which leaves the question whether changing the objective function this way is justified. It might lead to difficulties in updating the parameters of the objective functions, because of changes in the derivations of the parameter calculations. Kovács et al. (2021) found that the elastic net penalty can be implemented in the graphical lasso algorithm for sparse precision matrices by Friedman et al. (2008). They show that changing the penalisation of the target function can be implemented when small changes are made to updating the parameters. They describe two problems when changing the penalisation function. The first is finding derivations of the formula's to update column and rows in the matrices. The expressions of those updates can be derived similar to the derivation of the updates in Friedman et al. (2008). The second problem is the application of coordinate descent for the optimisation problem, which changes due to the elastic net penalisation. Kovács et al. (2021) found the same update procedure as described by Tibshirani et al. (2010), who derive the coordinate-wise-update formula for an elastic net problem. With both problems solved, their graphical elastic net (gelnet) algorithm is implemented to estimate the precision matrix.

4.3.1 Graphical elastic net with the correlation matrix

Kovács et al. (2021) mention that in practice the correlation matrix should be used in order to deal with variable scaling, because this becomes more important when using the elastic net penalty instead of the lasso penalty. Even though the return data is of the same scale, it is still interesting to investigate whether the correlation changes the performance of the gelnet algorithm.

To estimate the weight vector the estimated precision matrix based on the correlation matrix are re-scaled, such that the precision matrix is defined as follows:

$$\hat{\Sigma}^{-1} = Q\hat{C}^{-1}Q \tag{8}$$

where Q is a N x N diagonal matrix with $Q_{i,i} = \frac{1}{\hat{S}_{ii}}$, where $\hat{S}_{ii}$ is element (i,i) of $\hat{S}$ and Q_i element (i,i) of Q. $\hat{C}^{-1}$ is the estimated precision matrix by the gelnet algorithm with the sample correlation as input and $\hat{\Sigma}^{-1}$ is the estimated precision matrix used to calculate the weights.

4.4 Node-wise Regression

The limited amount of research in the field of the graphical elastic net, and especially the gelnet estimator by Kovács et al. (2021) leaves questions about the properties of the resulting estimator. Banerjee et al. (2008) proved for the glasso algorithm that the resulting estimator is positive definite and invertible, but this might not be true for the gelnet algorithm, due to the change in the penalised log-likelihood and the corresponding dual problem.

Node-wise regression, which was introduced by Meinshausen and Bühlmann (2006) as neighbourhood selection with the lasso, is a method for which adaptation to elastic net is more straightforward. Therefore this method is also implemented to evaluate if the elastic net penalty improves portfolio performance. Callot et al. (2021) used this method in the field of portfolio optimisation and showed how this node-wise regression can be applied to estimate a precision matrix. They found that the following regression:

$$r^*_{t,j} = (r^*_{t,-j})'\gamma_j + \eta_{t,j} \tag{9}$$

where $r^*_{t,j}$ is the demeaned excess return of stock j at time t, $r^*_{t,-j}$ is the vector of demeaned excess returns of all stocks except stock j at time t, γ_j is the vector of parameter estimates, and $\eta_{t,j}$ is the estimation error of the regression of stock j at time t. Note that Equation 9 is the same as Equation 1, but without α. Therefore the elements of the precision matrix are calculated using Equation 2.

Using Equation 9 leads to a similar optimisation as in Meinshausen and Bühlmann (2006), which is changed slightly to include elastic net penalisation as follows:

$$\hat{\gamma}_j = \underset{\gamma \in R^{p-1}}{argmin} \left(\frac{\|r^*_j - \boldsymbol{r^*}_{-j}\gamma\|_2^2}{T} + 2\lambda_j\alpha\|\gamma\|_1 + (1-\alpha)\lambda_j\|\gamma\|_2^2 \right). \tag{10}$$

Note that changing the penalisation function in this method does not change any of the derivations for estimating the precision matrix, because none of them are based on the objective function. This is an important advantage over the glasso estimator, for which the subgradients of the objective function are used to derive estimations. This is not needed for nodewise-regression since the relation between γ and Σ^{-1} only depends on the regression of $r^*_{t,j}$ on γ with elastic net penalty.

4.5 2 Stage Graphical Elastic Net

The 2 stage graphical elastic net (2S Gelnet) as used in Bernardini et al. (2021) is a combination of the graphical elastic net and node-wise regression methods. Node-wise regression is performed first to find a sparsity pattern for the precision matrix, because node-wise regression is the same as neighbourhood selection by Meinshausen and Bühlmann (2006), which is known to perform well for finding relations in graphs, which are the non-zero elements of the precision matrix. Meinshausen (2008) shows that glasso estimation can be inconsistent in certain situations where node-wise regression is still consistent. This gives reason to estimate the zero-elements of the precision matrix via node-wise regression before using the gelnet algorithm. Bernardini et al. (2021) show that using this constraint version of the graphical elastic net can lead to better estimation of the zero elements of a matrix. The graphical elastic net estimator is calculated with ρ defined as follows:

$$\rho_{i,j} = \begin{cases} \infty & \text{if } \hat{\gamma}_{i,j} = 0 \text{ or } \hat{\gamma}_{j,i} = 0 \\ \rho & \text{otherwise} \end{cases} \tag{11}$$

where $\rho_{i,j}$ is the (i,j) element of the N x N ρ matrix, which now sets some elements of the precision matrix to 0 before using the gelnet algorithm by penalising that element very heavily. $\hat{\gamma}_{i,j}$ is the i'th element of $\hat{\gamma}_j$, which is the vector of parameter estimates of the regression of the returns of the j'th asset on the other $N-1$ assets. In practice ∞ is set to a large number.

4.6 General Mean Variance Portfolio & Alternative Portfolio's

In general the optimal mean-variance portfolio is constructed by using the precision matrix to create optimal portfolio weights using the following expression:

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

where w is a N x 1 vector of portfolio weights, $\mathbf{1}_N$ is a N x 1 vector of ones, and $\hat{\Sigma}^{-1}$ is the estimator

of the precision matrix. The portfolio returns are calculated by using the different precision matrix estimators based on $r_{t-119}, r_{t-118}, ..., r_t$ to create portfolio weights for month $t+1$ and then multiply those weights with the asset returns in month $t+1$.

In this research every portfolio is constructed as the mean-variance portfolio. The glasso portfolio is denoted by Ψ_ρ, following notation by Goto and Xu (2015). The gelnet portfolio is denoted by $\Xi_{\rho,\alpha}$ and the node-wise regression portfolio is denoted by NR. The gelnet portfolio with the correlation matrix instead of the covariance matrix is denoted by $\Xi_{\rho,\alpha}$ (cor). The performance of these portfolio's is compared to the equal weighted, sample covariance, Jagannathan-Ma and Ledoit-Wolf portfolio's.

The Equal Weighted (1/N) portfolio, for which every weight $w_j = 1/N$, where w is the N x 1 weight vector, is easy to calculate, since the weights do not have to be computed via the precision matrix. Despite it's simplicity DeMiguel et al. (2009) found that none of the more complicated portfolio's in their research consistently outperformed the 1/N portfolio. The Sample Covariance ($\hat{S}^{-1}$) portfolio is calculated by using the inverse of the sample covariance matrix as precision matrix estimator. The Jagannathan-Ma (JM) portfolio by Jagannathan and Ma (2003) is similar to the $\hat{S}^{-1}$ portfolio, because it uses the sample covariance matrix, but for this portfolio the weights are calculated by optimising the following function: $\min_w w'Sw$ with 2 constraints: $\sum_{i=1}^{N} w_i = 1$ and $w_i \geq 0$, which leads to a portfolio without short positions. Even though it seems very similar to the sample covariance portfolio, this portfolio has proven to reduce the risk of the portfolio significantly compared to the sample covariance portfolio. The Ledoit-Wolf (LW) portfolio is calculated by estimating the precision matrix using the shrinkage method proposed by Ledoit and Wolf (2004). They use a weighted average of the sample covariance matrix and the constant correlation matrix to shrink the covariance estimator towards the constant correlation matrix.

4.7 Parameter Optimisation

For the glasso, gelnet and nodewise regression methods there are certain parameters that need to be optimised in order to create an estimate of the precision matrix. The training period, which contains the first 240 months of data (July 1963 - June 1983), is used to optimise the parameters. 120 precision matrices for period t between July 1973 and June 1983 are estimated based on $r_{t-120}, r_{t-119}, ..., r_{t-1}$. When the optimal parameters based on this training period are found, they are used for out-of-sample prediction between July 1983 and December 2010. The precision matrices for period t in the out-of-sample period are also estimated using $r_{t-120}, r_{t-119}, ..., r_{t-1}$.

4.7.1 Predictive likelihood

The parameters of the gelnet and glasso estimators are optimised by performing a grid search for $0 \leq \rho \leq 10$ and $0 \leq \alpha \leq 1$ with increments of 0.1 to maximise the predictive Gaussian likelihood function, following Goto and Xu (2015), which is defined as follows:

$$L(\hat{\Sigma}^{-1}) = \frac{1}{T_f} \sum_{t=1}^{T_f} l_t(\hat{\Sigma}^{-1}) \tag{13}$$

where $l_t(\hat{\Sigma}^{-1}) = \ln(\det(\hat{\Sigma}_{t-1}^{-1})) - \tilde{R}_t' \hat{\Sigma}_{t-1}^{-1} \tilde{R}_t$. T_f is the number of observations in the estimation sample, $\hat{\Sigma}_{t-1}^{-1}$ is the precision matrix estimated at period $t-1$ and $\tilde{R}_t$ is the demeaned return vector $\tilde{R}_t = R_t - \frac{1}{T_f} \sum_{i=1}^{T_f} R_i$, which is the estimation sample mean subtracted from the return vector.

4.7.2 Generalized Information Criterion

The parameters ρ and α of the Node-wise Regression estimator are optimised by using a different information criterion instead of the predictive likelihood. Y. Fan and Tang (2013) found that their Generalized Information Criterion (GIC) finds the true model with probability approaching one if it is used as information criterion for Nodewise-Regression. This criterion is defined as follows:

$$GIC(\lambda_j) = \ln(\hat{\sigma}_{\lambda_j}^2) + \sum_{i=1}^{N} I\{\hat{\gamma}_{i,j}(\lambda_j) \neq 0\} \frac{\ln(N)}{T} \ln(\ln(T)) \tag{14}$$

where $\hat{\sigma}_{\lambda_j}^2 = \|r_j - r_{-j}\hat{\gamma}_j\|_2^2$, which is the residual variance of the regression of the return of the j'th asset (r_j) on the return of all assets except the j'th (r_{-j}). $I\{\hat{\gamma}_{i,j}(\lambda_j) \neq 0\}$ is an indicator function, which is 1 if $\hat{\gamma}_{i,j}(\lambda_j) \neq 0$ and 0 otherwise, where $\hat{\gamma}_{i,j}(\lambda_j)$ is element i of parameter vector $\hat{\gamma}_j$, which is the vector of parameter estimates of the regression of the return of the j'th asset on the other $N-1$ assets. T is the number of periods in the estimation sample and N is the number of assets.

The GIC allows node-wise regression to use a different optimal penalisation for every asset, because it optimises every regression separately. To allow this difference, ρ and α are optimised every month for node-wise regression, instead of set to a fixed value optimised in the training period. This could be advantageous compared to the predictive likelihood if the amount of penalisation is different over time or different between different assets.

5 Results

All results are obtained via R programs in R 4.0.5. Next to the basic R functions the following packages are used: glasso, GLassoElnetFast, matlib, glmnet, matrixcalc and tseries. For the Ledoit-Wolf estimator by Ledoit and Wolf (2004) and the circulair bootstrap by Ledoit and Wolf (2007) the code from Ledoit's UZH page is used (https://www.econ.uzh.ch/en/people/faculty/wolf/publications.html). For the optimal block sizes of the bootstrap methods, the code by Patton et al. (2009) is used, which is available on http://www.math.ucsd.edu/~politis/SOFT/PPW/ppw.R.

5.1 Optimal parameter choices

The optimal parameters play an important role in order to create portfolios based on the glasso, gelnet or node-wise regression estimators. Table 2 shows the optimal parameter choices for each data set. Interestingly, α is on average close to 1 (0.948) for the NR estimator for every data set. This means that node-wise regression does not really benefit from using the elastic net, because when α is optimised for $0 \leq \alpha \leq 1$, $\alpha \approx 1$ performs best, which is almost equal to lasso regularisation. Note that the NR portfolio has low sparsity, which implies that ρ is small for the NR portfolio.

Table 2: Optimal Parameters and Sparsity

Note that for the glasso estimator Ψ_ρ, $\alpha = 1$ for every data set, because it does not allow for different α. The reported α and ρ for NR are averages over every regression performed in the out-of-sample period. The reported sparsity is the percentage of zero-elements in the precision matrices of the out-of-sample period.

Dataset	Parameter	$\hat{\Psi}_\rho$	$\hat{\Xi}_{\rho,\alpha}$	NR	$\hat{\Xi}_{\rho,\alpha}$ (cor)
SZBM100	ρ	1	2.5	3.2	0.1
	α	1	0.4	0.9	0.1
	Sparsity (%)	49.1	49.0	11.6	53.2
IND48	ρ	1.3	2.1	4.4	0.1
	α	1	0.6	0.9	0.1
	Sparsity (%)	42.7	42.8	14.1	50.5
SZBM100+IND48	ρ	1.3	1.6	4.6	0.1
	α	1	0.8	0.9	0.1
	Sparsity (%)	46.2	46.2	8.3	52.9

Another interesting result from Table 2 is that $\rho_{\hat{\Xi}} \times \alpha_{\hat{\Xi}} \approx \rho_\Psi$, where $\alpha_{\hat{\Xi}}$ and $\rho_{\hat{\Xi}}$ are ρ and α from Equation 7. This leads to a very similar penalisation function, which leads to similar portfolio

weights for $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$. An explanation might be that the absolute penalty is always larger than the quadratic penalty, therefore the algorithm only has to minimise the absolute penalty. This can be the case if the elements of the precision matrix are smaller than one. The optimal parameters for the $\hat{\Xi}_{\rho,\alpha}$ (cor) portfolio are all 0.1, which is the smallest amount of regularisation. For the SZBM100+IND48 data set the predictive likelihood was $-\infty$, which suggests that a smaller increment should be used, because for the other methods $-\infty$ only occurs when α or ρ is too large.

Some of the values for ρ or α for $\hat{\Xi}_{\rho,\alpha}$ lead to predictive likelihoods that caused problems during optimisation, because the determinant of the precision matrix is equal to zero. This means that the predictive likelihood could not be calculated, because it would lead to $\log(0)$. Furthermore this means that the precision matrix is not invertible. Banerjee et al. (2008) proved that if you put an invertible matrix in the glasso algorithm, that the output would be invertible too, but this does not seem to hold for graphical elastic net. This problem should be looked in to in future research.

For the $\hat{\Psi}_\rho$ portfolio a different optimal ρ is found than reported by Goto and Xu (2015) for the SZBM100 and the SZBM100+IND48 data set. This difference might be due to a change in the SZBM100 data set in August 2020. French's data library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html) reports that there have been made adjustments to book equity, which affects portfolios that are based on book-to-market equity. The SZBM100 and SZBM100+IND48 data sets depend on those adjustments, which explains the differences. This means only the optimal parameters for the IND48 data set should be the same, which is the case.

Finally it is interesting that calculations for the glasso and gelnet estimators with the correlation matrix are very fast compared to the same method with the covariance matrix. This might be beneficial if the optimal parameters are re-estimated multiple times, for every month for example. This may in turn lead to a lower return variance, due to the better optimised parameters.

5.2 Return Variance

An important aspect of the different inverse covariance estimators is the variance of the returns of minimum variance portfolios based on the different estimators. The return variance is calculated by taking the variance of out-of-sample returns of the minimum variance portfolios.

Table 3 contains the return variances and shows that $\hat{\Xi}_{\rho,\alpha}$ has the lowest return variance for every data set, but it is close to the return variance of $\hat{\Psi}_\rho$. The return variance of $\hat{\Xi}_{\rho,\alpha}$ (cor) is close to $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ for all data sets, with lower variance for the IND48 data set and higher variance for the other data sets.

Table 3: Return Variance of Out-of-Sample Returns

The ranking of best (1), which is lowest out-of-sample return variance or risk, to worst (8), which is highest return variance or risk, is reported below the return variance of each portfolio. $N > T$ means that this portfolio could not be constructed, due to the non-invertable sample covariance matrix.

Dataset	$\hat{\Psi}_\rho$	$\hat{S}^{-1}$	1/N	JM	LW	$\hat{\Xi}_{\rho,\alpha}$	NR	$\hat{\Xi}_{\rho,\alpha}$(cor)
SZBM100	13.779	59.609	25.834	18.407	18.879	13.736	22.296	14.584
	(2)	(8)	(7)	(4)	(5)	(1)	(6)	(3)
IND48	12.355	17.315	22.796	13.326	13.012	12.351	17.248	12.231
	(3)	(6)	(8)	(5)	(4)	(2)	(7)	(1)
SZBM100+IND48	10.563	$N > T$	24.327	12.755	12.023	10.546	21.102	11.953
	(2)		(7)	(5)	(4)	(1)	(6)	(3)

The results of all portfolios are close to the findings of Goto and Xu (2015), except for the JM portfolio, which has a lower variance. This difference might be due to the change in the French Data Library mentioned in Section 5.1. The NR portfolio does not perform as well. It has return variance almost always closer to the 1/N portfolio than to the $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ portfolios. This shows that the NR portfolio is riskier than the $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ portfolios.

Table 4: Stationary Bootstrap of Return Variance compared to Glasso

Table 4 reports the test statistic of the stationary bootstrap by Politis and Romano (1994), with optimal block sizes from Politis and White (2004). All portfolios are compared to $\hat{\Psi}_\rho$ as follows: $\sigma_X - \sigma_{\hat{\Psi}_\rho}$, where σ_X is the standard deviation of returns of portfolio X. ***, ** and * denote a significant difference at a significance level of 1%, 5% and 10% respectively. $N > T$ means that this portfolio could not be constructed, due to the non-invertable sample covariance matrix.

Dataset	$\hat{S}^{-1}$	1/N	JM	LW	$\hat{\Xi}_{\rho,\alpha}$	NR	$\hat{\Xi}_{\rho,\alpha}$ (cor)
SZBM100	4.000***	1.358***	0.580*	0.623*	-0.003	1.000**	0.117
IND48	0.645**	1.240***	0.135	0.094	-0.000	0.620*	-0.023
SZBM100+IND48	$N > T$	1.679***	0.314	0.225	-0.009	1.336***	0.194

The difference between the return variances is evaluated by testing the null hypothesis of equal return variance between the return variance of the glasso estimator and the $\hat{S}^{-1}$, 1/N, JM, LW and NR estimators. Table 4 shows the results of this test, which is calculated using the stationary

bootstrap by Politis and Romano (1994) with optimal block sizes from Politis and White (2004). Note that Table 4 shows that there is no significant difference in return variance between the $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ estimator on a $\leq 10\%$ significance level. Therefore changing from lasso to elastic net penalty seems to not improve the optimal mean-variance portfolio based on out-of-sample return variance.

Another interesting result is that the JM estimator performs better than reported by Goto and Xu (2015). They find significant differences for every data set and Table 4 shows only a significant difference for the SZBM100 data set. This result shows that simple portfolio strategies, such as the nonnegative portfolio weights, can still lead to a low risk portfolio.

5.3 Weight Distribution

Because of the small difference in return variance between $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$, the weight distribution of the portfolios are especially of interest to see if the methods result in different portfolios. Table 2 shows that penalisation is approximately the same for $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$, which raises the question if these portfolios are any different. The weight distribution of $\hat{\Xi}_{\rho,\alpha}$ should be different, because elastic net penalisation should result in portfolio weights closer to zero compared to lasso penalisation. Table 5 shows however that the weight distribution of $\hat{\Xi}_{\rho,\alpha}$ is close to that of $\hat{\Psi}_\rho$. The minimum, 1%, 5%, 95% quantiles and maximum weights are closer to zero for the SZBM100 data set as expected, but only by a small amount. The Herfindahl index, defined as $\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} w_{i,t}$ where $w_{i,t}$ is the weight of asset i in period t, confirms that $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ are close in terms of weights distribution. For the IND48 data set $\hat{\Xi}_{\rho,\alpha}$ has larger absolute weights. This result might be explained by the magnitude of the elements of the precision matrix. If they are smaller than one, lasso penalisation is larger, which results in smaller portfolio weights.

$\hat{\Xi}_{\rho,\alpha}$ (cor) has weights closer to zero compared to $\hat{\Xi}_{\rho,\alpha}$ and $\hat{\Psi}_\rho$. This is in line with the earlier note that the penalisation might be too large. If penalisation was smaller, larger portfolio weights would occur. The lower portfolio weights confirm that $\hat{\Xi}_{\rho,\alpha}$ (cor) seems too heavily penalised, because it has higher variance for two data sets, which could be reduced by allowing larger portfolio weights.

NR has noticeably smaller portfolio weights, except for IND48 min. and max., and is able to create a portfolio that seems to be better than the equal weighted portfolio, but still has very small portfolio weights. The Herfindahl index confirms that NR has small weights. It seems that the optimisation of ρ and α should be adjusted for SZBM100 and SZBM100+IND48 to allow larger weights. This way risk reduction can be realised by allowing larger weights for low risk assets. The weight distributions of the JM and LW portfolios are reported in Appendix A.2 Table A1.

Table 5: Weight Distributions

$N > T$ means that this portfolio could not be constructed, due to the non-invertable sample covariance matrix. 1%, 5%, 95% and 99% are the quantiles of all weights of the out-of-sample portfolios.

Data set	Portfolio	Min.	1%	5%	95%	99%	Max.	Herfindahl
SZBM100	$\hat{\Psi}_\rho$	−0.261	−0.151	−0.108	0.149	0.214	0.313	0.615
	$\hat{S}^{-1}$	−1.934	−0.748	−0.501	0.533	0.836	1.852	10.124
	$\hat{\Xi}_{\rho,\alpha}$	−0.261	−0.150	−0.108	0.148	0.214	0.310	0.607
	NR	−0.037	−0.009	−0.002	0.029	0.058	0.122	0.025
	$\hat{\Xi}_{\rho,\alpha}$ (cor)	−0.116	−0.077	−0.055	0.112	0.162	0.290	0.277
IND48	$\hat{\Psi}_\rho$	−0.233	−0.151	−0.106	0.161	0.382	0.511	0.414
	$\hat{S}^{-1}$	−0.736	−0.377	−0.202	0.251	0.529	0.843	1.088
	$\hat{\Xi}_{\rho,\alpha}$	−0.233	−0.153	−0.106	0.161	0.382	0.512	0.416
	NR	−0.504	−0.043	−0.010	0.086	0.159	0.302	0.082
	$\hat{\Xi}_{\rho,\alpha}$ (cor)	−0.124	−0.092	−0.061	0.166	0.495	0.789	0.441
SZBM100+IND48	$\hat{\Psi}_\rho$	−0.186	−0.108	−0.069	0.093	0.146	0.287	0.380
	$\hat{S}^{-1}$				$N > T$			
	$\hat{\Xi}_{\rho,\alpha}$	−0.185	−0.108	−0.069	0.093	0.146	0.289	0.381
	NR	−0.037	−0.004	−0.001	0.021	0.039	0.115	0.017
	$\hat{\Xi}_{\rho,\alpha}$ (cor)	−0.101	−0.057	−0.037	0.077	0.123	0.328	0.205

5.4 Sharpe Ratio

Table 6 shows the Sharpe ratio of all different estimators. Note that $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ are again the best for the SZBM100 data set and the SZBM100+IND48 data set. It is also interesting to see that $\hat{\Xi}_{\rho,\alpha}$ does have a significantly higher Sharpe ratio than $\hat{\Psi}_\rho$ for the SZBM100 data set on a 10% significance level. This suggests that using the elastic net penalty might lead to a more efficient estimator, because of the higher reward to risk ratio. $\hat{\Xi}_{\rho,\alpha}$ (cor) is outperformed by both $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$, which is in line with the results for the return variance. Note that $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ are not the best choice for the IND48 data set, because the NR portfolio has a higher Sharpe ratio.

It is quite surprising that $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ are outperformed by the NR , 1/N and the JM portfolio for the IND48 data set. The difference between data sets might be explained by the $\frac{N}{T}$ ratio, because this ratio is higher for the SZBM100 and SZBM100+IND48 data sets. $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ are designed to create more sparse portfolios to reduce estimation errors for data sets that suffer from high $\frac{N}{T}$ ratio or high correlations. This might explain why these portfolios perform significantly better for

Table 6: Out-of-sample Sharpe Ratio

Data set	$\hat{\Psi}_\rho$	$\hat{S}^{-1}$	1/N	JM	LW	$\hat{\Xi}_{\rho,\alpha}$	NR	$\hat{\Xi}_{\rho,\alpha}$ (cor)
SZBM100	0.292	0.171	0.135	0.173	0.241	0.294	0.164	0.238
IND48	0.128	0.075	0.131	0.133	0.121	0.128	0.143	0.132
SZBM100+IND48	0.288	$N > T$	0.135	0.171	0.285	0.288	0.159	0.237

the SZBM100 and SZBM100+IND48 data sets. On the other hand, when multi-collinearity is not a big problem, NR performs equally, or might even perform better in terms of Sharpe ratio.

Table 7 shows the statistics of the circular bootstrap by Ledoit and Wolf (2007) with optimal block sizes. It shows that the differences between $\hat{\Psi}_\rho$ and the other portfolios are not significantly different from each other for the IND48 data set, except for the $\hat{S}^{-1}$ portfolio. This means that for the IND48 data set all estimation methods perform about the same based on Sharpe ratio. Because of the significant differences for the other data sets $\hat{\Xi}_{\rho,\alpha}$ seems to be the most efficient in terms of Sharpe ratio, since it significantly outperforms $\hat{\Psi}_\rho$ for the SZBM100 data set.

Table 7: Circulair Bootstrap Test Statistics compared with Glasso

The reported statistic is the results of the circulair bootstrap by Ledoit and Wolf (2007), where every portfolio is compared to $\hat{\Psi}_\rho$ as follows: $Sharpe_{\hat{\Psi}_\rho} - Sharpe_X$, where $Sharpe_X$ is the Sharpe ratio of portfolio X. ***, ** and * denote a significant difference at a significance level of 1%, 5% and 10% respectively. $N > T$ means that this portfolio could not be constructed, due to the non-invertable sample covariance matrix.

Data set	$\hat{S}^{-1}$	1/N	JM	LW	$\hat{\Xi}_{\rho,\alpha}$	NR	$\hat{\Xi}_{\rho,\alpha}$ (cor)
SZBM100	0.122*	0.157***	0.119***	0.052	-0.001*	0.129***	0.054**
IND48	0.053**	-0.003	-0.005	0.007	0.001	-0.015	-0.003
SZBM100+IND48	$N > T$	0.152***	0.117***	0.002	0.000	0.128**	0.051*

5.5 Turnover

In this section the stability in terms of weights over time is evaluated. Due to the transaction costs involved with portfolio changes, a stable low risk portfolio is preferred over one that changes weights extremely after a period. Table 8 reports the turnover for every portfolio for each data set. Lower turnover means less transaction costs, so the 1/N portfolio has the best turnover, followed by JM. $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ have higher turnover than NR, which was expected, due to the larger portfolio weights in the weight distribution, which results in larger turnover when the optimal weights are changed.

Table 8: Out-of-sample Portfolio Turnover

Data set	$\hat{\Psi}_\rho$	$\hat{S}^{-1}$	1/N	JM	LW	$\hat{\Xi}_{\rho,\alpha}$	NR	$\hat{\Xi}_{\rho,\alpha}$(cor)
SZBM100	0.641	7.093	0.024	0.132	0.585	0.632	0.193	0.339
IND48	0.299	0.775	0.033	0.077	0.329	0.300	0.258	0.203
SZBM100+IND48	0.611	$N > T$	0.027	0.090	0.760	0.619	0.173	0.346

$\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$ are close in terms of turnover and $\hat{\Xi}_{\rho,\alpha}$ (cor) in between NR and $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$, which is in line with the small differences in weight distribution.

5.6 Economic Gain

The economic gain of every portfolio is calculated to combine the results found for return variance, Sharpe Ratio and turnover. The economic gain is evaluated by calculating the transaction cost adjusted certainty equivalent excess return (TCACER) with $\gamma = 5$, following Goto and Xu (2015). γ is the risk-aversion coefficient. The TCACER by Goto and Xu (2015) is defined as follows:

$$TCACER = \hat{\mu}_q - \frac{\gamma}{2}\hat{\sigma}_q^2 - T\_COST_q \tag{15}$$

where $\hat{\mu}_q$ and $\hat{\sigma}_q^2$ are the annualised mean return and return variance respectively and T_COST is the annualised transaction cost. The TCACER can be interpreted as the percentage of risk free return for which an investor is indifferent between the risk free return and a portfolio of risky assets.

Table 9 reports the TCACER of every portfolio. Note that the results are different than reported by Goto and Xu (2015), especially for the 1/N, JM and LW portfolios. Ψ_ρ and $\hat{\Xi}_{\rho,\alpha}$ are very similar and the best choices based on the SZBM100 and SZBM10+IND48 data sets. For the IND48 data set they do perform worse, which was also the case for the Sharpe ratio. Interestingly $\hat{\Xi}_{\rho,\alpha}$ (cor) and the NR portfolio also perform well, because they have positive TCACER for each data set. On average $\hat{\Xi}_{\rho,\alpha}$ (cor) outperforms NR and is almost equal to $\hat{\Psi}_\rho$ and $\hat{\Xi}_{\rho,\alpha}$. $\hat{\Xi}_{\rho,\alpha}$ (cor) therefore may be more robust, since it always has positive economic gain for each of the data sets.

Table 9: Transaction Cost Adjusted Certainty Equivalent Excess Returns (TCACER's)

Data set	$\hat{\Psi}_\rho$	$\hat{S}^{-1}$	1 /N	JM	LW	$\hat{\Xi}_{\rho,\alpha}$	NR	$\hat{\Xi}_{\rho,\alpha}$(cor)
SZBM100	5.040	−44.627	0.362	2.616	3.384	5.143	1.419	4.512
IND48	−0.086	−6.093	0.457	1.364	−0.640	−0.111	0.408	0.638
SZBM+IND48	4.386	$N > T$	0.547	2.955	3.706	4.344	1.418	4.171

5.7 2 Stage Graphical Elastic Net

The 2S Gelnet estimator is obtained by optimising ρ based on GIC and α based on predictive likelihood for the NR estimator and then maximising the predictive likelihood of the gelnet algorithm with certain elements of the precision matrix constraint to be zero. The difference with the normal NR estimator is made, such that α_{NR} is chosen to fit the combined method (node-wise regression and gelnet) best, instead of the NR estimator only. Note that therefore α_{NR} is the same for every regression. Optimising α shows that $\alpha_{NR} = 0$ is optimal for the 2S Gelnet algorithm. This means that almost no constraints are set on the estimation of the gelnet precision matrix. Therefore the results of the 2S Gelnet portfolio are close to the results of $\hat{\Xi}_{\rho,\alpha}$, which is confirmed by Table 10, where Column 1 and 3 are very close. Column 2 shows the return variance when $\alpha > 0$, which shows that it leads to higher return variance. This was not entirely expected, because NR is known to find relations between variables well, which should lead to a better portfolio. On the other hand, putting a constraint usually only restricts the outcome, which is the case here. The other statistics, such as Sharpe ratio, weight distribution and economic gain are all similar to $\hat{\Xi}_{\rho,\alpha}$, because 2S Gelnet leads to almost the same weights as $\hat{\Xi}_{\rho,\alpha}$.

Table 10: Return Variance 2S Gelnet

Data set	2S Gelnet	2S Gelnet $\alpha = 0.1$	$\hat{\Xi}_{\rho,\alpha}$
SZBM100	13.748	14.182	13.736
IND48	12.354	12.398	12.351
SZBM100+IND48	10.581	10.609	10.546

6 Conclusion

To conclude this research, the performance of different precision matrix estimators is compared with the glasso estimator by Goto and Xu (2015). This is done by first answering the first research question: *"Can the graphical elastic net improve the performance of the improved mean-variance estimator?"*. The gelnet portfolio has proven to perform a little better than the graphical lasso portfolio, with a significant improvement of the Sharpe ratio for one of the data sets. The gelnet correlation estimator performs worse, but has on average equal economic gain. All things considered, the answer to the first research question is yes, but there is not always a significant improvement.

Node-wise regression with the elastic net penalty performs significantly worse than both the

glasso and gelnet portfolios based on return variance and Sharpe Ratio when the number of stocks is close to the number of observations. When this is not the case, node-wise regression has higher return variance, but equal Sharpe ratio. The optimal α for node-wise regression, which was on average close to 1, implies that changing from lasso to elastic net penalty does not improve performance of node-wise regression. The second research question: *"Can node-wise regression with elastic net penalty improve performance of the improved mean-variance estimator?"* can therefore be answered with no, because the glasso portfolio performs better when multi-collinearity problems arise.

The 2 stage graphical net estimator reduced to the gelnet estimator when the parameters are optimised, which means that combining node-wise regression with graphical elastic net does not improve the glasso estimator. So the last research question: *"Can 2 stage graphical elastic net improve performance of the improved mean-variance estimator?"* is answered with no, because the optimal form reduces to gelnet estimation, which is faster and should be preferred.

The answer to the main research question: *"Can the performance of the improved mean-variance estimator be improved by using an elastic net penalty?"* is yes, by using the gelnet algorithm, but there is not always a significant difference. The gelnet estimator has shown that it might lead to significantly better Sharpe ratio and that it at least performs as well as the glasso estimator for different data sets. Other estimators, such as the node-wise regression estimator seems to perform well when the number of stocks is sufficiently smaller than the number of observations.

In practice the glasso estimator or the gelnet correlation matrix estimator might still be preferred over the gelnet estimator, due to the huge computational benefit when estimating matrices for high-dimensional data. The gelnet correlation matrix estimator especially shows some potential, because it can be improved by lowering the regularisation parameters and is computationally faster, allowing for re-optimising the parameters more often. The economic gain of the gelnet correlation matrix portfolio was about equal to the glasso and gelnet portfolios on average, but it was always positive, which was not the case for the glasso and gelnet portfolios. This suggests that it may be more robust to different data.

A Appendix

A.1 Details on the Graphical Lasso calculations

To maximise Equation 4 Friedman et al. (2008) use a graphical lasso (glasso) algorithm to estimate the parameters of the QML maximisation. They make use of the dual problem of the equation found by Banerjee et al. (2008), which is defined following notation of Friedman et al. (2008) as follows:

$$\min_{\beta}\left\{\frac{1}{2}\|W_{11}^{\frac{1}{2}}\beta - b\|^2 + \rho\|\beta\|_1\right\} \tag{16}$$

where $b = W_{11}^{-\frac{1}{2}}\hat{s}_{12}$, $\beta = W_{11}^{-1}w_{12}$, where W_{ij}, w_{ij} and $\hat{s}_{ij}$ are partitions of W and $\hat{S}$. W is the estimate of Σ and $\hat{S}$ is the sample covariance estimator. Banerjee et al. (2008) show that if this algorithm starts with a positive definite matrix, that the end result will be positive definite too. This is an important feature, since a positive definite covariance matrix is needed in order to invert it. W and $\hat{S}$ are partitioned as follows:

$$W = \begin{bmatrix} W_{11} & w_{12} \\ w_{12}' & w_{22} \end{bmatrix}, \quad \hat{S} = \begin{bmatrix} \hat{S}_{11} & \hat{s}_{12} \\ \hat{s}_{12}' & \hat{s}_{22} \end{bmatrix} \tag{17}$$

where w_{ij} and $\hat{s}_{ij}$ are the last columns/rows of W and $\hat{S}$ respectively. Permuting this last column and row, allows W to be updated by changing the last column and row and then re-estimating W. This is done by using the graphical lasso algorithm from Friedman et al. (2008):

1. Start with $W = S + \rho I$.
2. For each j = 1,2,...,p,1,2,...,p,..., solve the lasso problem (Equation 16), which takes the inner products W_{11} and $\hat{S}$ as input. This will give an estimate $\hat{\beta}$ which we use to update $w_{12} = W_{11}\hat{\beta}$.
3. Repeat until the problem has converged

They also show that $\hat{\Sigma}^{-1}$ can easily be calculated due to the relation between W and Σ^{-1}, $W\Sigma^{-1} = I$, from which Σ_{12}^{-1} and Σ_{22}^{-1} are derived. Here Σ_{12}^{-1} and Σ_{22}^{-1} form the last column of Σ^{-1}. Because we find W by permuting the last column, such that every column will be the last column at some point, we can calculate every column of Σ^{-1} quite easily.

A.2 Weight Distribution JM and LW Portfolio's

Table A1 contains the weighs distributions of the remaining two portfolio's: the JM and LW portfolio's.

Table A1: Weight Distribution JM and LW portfolio's

1%, 5%, 95% and 99% are the quantiles of all weights of the out-of-sample portfolio's.

Data set	Portfolio	Min.	1%	5%	95%	99%	Max.	Herfindahl
SZBM100	JM	0.000	0.000	0.000	0.077	0.234	0.468	0.185
	LW	−0.200	−0.132	−0.092	0.143	0.206	0.357	0.545
IND48	JM	0.000	0.000	0.000	0.110	0.527	0.718	0.307
	LW	−0.244	−0.148	−0.104	0.177	0.492	0.657	0.524
SZBM100+IND48	JM	0.000	0.000	0.000	0.021	0.150	0.696	0.288
	LW	−0.180	−0.110	−0.077	0.111	0.178	0.401	0.507

References

Banerjee, O., Ghaoui, L. E., & d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine Learning Research*, *9*(15), 485–516. http://jmlr.org/papers/v9/banerjee08a.html

Bernardini, D., Paterlini, S., & Taufer, E. (2021). New estimation approaches for graphical models with elastic net penalty.

Callot, L., Caner, M., Önder, A. Ö., & Ulaşan, E. (2021). A nodewise regression approach to estimating large portfolios. *Journal of Business & Economic Statistics*, *39*(2), 520–531. https://doi.org/10.1080/07350015.2019.1683018

DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, *55*(5), 798–812. http://www.jstor.org/stable/40539189

Demiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, *22*. https://doi.org/10.1093/rfs/hhm075

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/https://doi.org/10.1016/0304-405X(93)90023-5

Fan, J., Fan, Y., & Lv, J. (2008). High dimensional covariance matrix estimation using a factor model [Econometric modelling in finance and risk management: An overview]. *Journal of Econometrics*, *147*(1), 186–197. https://doi.org/https://doi.org/10.1016/j.jeconom.2008.09.017

Fan, Y., & Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *75*(3), 531–552. http://www.jstor.org/stable/24772736

Friedman, J., Hastie, T., Höfling, H., & Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, *1*. https://doi.org/10.1214/07-AOAS131

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, *9*, 432–41. https://doi.org/10.1093/biostatistics/kxm045

Goto, S., & Xu, Y. (2015). Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis*, *50*(6), 1415–1441. https://doi.org/10.1017/S0022109015000526

Ho, M., Sun, Z., & Xin, J. (2015). Weighted elastic net penalized mean-variance portfolio design and computation. *SIAM Journal on Financial Mathematics*, *6*. https://doi.org/10.1137/15M1007872

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hwang, I., Xu, S., & In, F. (2017). Naive versus optimal diversification: Tail risk and performance. *European Journal of Operational Research*, *265*. https://doi.org/10.1016/j.ejor.2017.07.066

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, *58*(4), 1651–1683. https://doi.org/https://doi.org/10.1111/1540-6261.00580

Kovács, S., Ruckstuhl, T., Obrist, H., & Bühlmann, P. (2021). Graphical elastic net and target matrices: Fast algorithms and software for sparse precision matrix estimation.

Ledoit, O., & Wolf, M. (2007). Robust performance hypothesis tests with the sharpe ratio. *Journal of Empirical Finance*, *15*, 850–859. https://doi.org/10.1016/j.jempfin.2008.03.002

Ledoit, O., & Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, *30*(4), 110–119. https://doi.org/10.3905/jpm.2004.110

Li, J. (2015). Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business Economic Statistics*, *33*(3), 381–392. http://www.jstor.org/stable/43701549

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91. http://www.jstor.org/stable/2975974

Meinshausen, N. (2008). A note on the lasso for gaussian graphical model selection. *Statistics Probability Letters*, *78*(7), 880–884. https://doi.org/https://doi.org/10.1016/j.spl.2007.09.014

Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, *34*(3), 1436–1462. https://doi.org/10.1214/009053606000000281

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, *45*(1), 31–42. http://www.jstor.org/stable/4479185

Nard, G. D., Ledoit, O., & Wolf, M. (2018). Factor models for portfolio selection in large dimensions: the good, the better and the ugly. (290). https://ideas.repec.org/p/zur/econwp/290.html

Patton, A., Politis, D. N., & White, H. (2009). Correction to "automatic block-length selection for the dependent bootstrap" by d. politis and h. white. *Econometric Reviews*, *28*(4), 372–375. https://doi.org/10.1080/07474930802459016

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*(428), 1303–1313. https://doi.org/10.1080/01621459.1994.10476870

Politis, D. N., & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, *23*(1), 53–70. https://doi.org/10.1081/ETC-120028836

Stevens, G. (1995). On the inverse of the covariance matrix in portfolio analysis. *International Finance Discussion Paper*, *1995*, 1–15. https://doi.org/10.17016/IFDP.1995.528

Tibshirani, R., Hastie, T., & Friedman, J. (2010). Regularized paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*. https://doi.org/10.1163/ej.9789004178922.i-328.7

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. http://www.jstor.org/stable/2346178

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net (vol b 67, pg 301, 2005). *Journal of the Royal Statistical Society Series B*, *67*, 768–768. https://doi.org/10.1111/j.1467-9868.2005.00527.x