Journal of Quantitative Finance and Economics
Volume 4; Number 1; 2022; pp : 115-132
https://doi.org/10.46791/jqfe.2022.v04i01.08

ARF INDIA
Academic Open Access Publishing
www. arfjournals. com

Portfolio Selection with Sparse Inverse Covariance Matrices

Jinjin Hu

Department of Statistics, School of Management, Fudan University, Shanghai 200433, China.
E-mail: hujj@fudan.edu.cn

ARTICLEINFO

Received: 8 February 2022
Revised: 30 March 2022
Accepted: 27 April 2022
Online: 30 May 2022

To cite this paper:

Jinjin Hu (2022). Portfolio Selection with Sparse Inverse Covariance Matrices. *Journal of Quantitative Finance and Economics.* 4(1), 115-132.

Abstract: *The estimation risk of the covariance matrix in portfolio selection often leads to poor out-of-sample performance, especially when the number of assets is large compared to the observation period. Shrinking the covariance matrix to a particular target which often based on financial theory has been used to reduce the standard errors of the estimates. In this paper, we propose to shrink the off-diagonal elements of the inverse covariance matrix to zeros in the estimation of the covariance matrix, the sparsity structure of the inverse covariance matrix imply that some of the assets are conditionally independent. Simulation study and empirical data analysis show that the new strategy based on sparse inverse covariance constraints often perform better than strategies with existing shrinkage estimates in terms of out-of-sample Sharpe ratio, variances and turnovers. In addition, the algorithm and the convergence rate of arriving the sparsity structure in the inverse covariance matrix has been provided.*

Keywords: *Portfolio selection, Estimation risk, Inverse covariance, Sparsity, Conditional independence.*

JEL classification: *G11*

1. Introduction

In the Mean Variance (MV) portfolio optimization model pioneered by Markowitz (1952), the mean and covariances of the asset returns are unknown parameters. Traditional econometric method is to estimate those parameters from historical data, then plug the estimates into the MV framework to derive the optimal portfolio weights. It has been found that replacing true parameters by their sample estimates may result in poor out-of-sample performance of Markowitz portfolio (for example, see Demiguel *et al.*. (2009b); Brandt (2009); Michaud (1989); Jobson and Korkie (1980)). Merton (1980) pointed out that it is more difficult to estimate means than estimating covariances of asset returns. Jagannathan and Ma (2003) reported that the portfolios are often efficient if one minimizes the portfolio variance by ignoring the constraint on means of the assets. Kourtis, Dotsis and Markellos (2012) reaffirmed the appeal of Global Minimum Variance (GMV) portfolio which only based on covariance matrix estimates to the

Mean Variance (MV) portfolio by analyzing the opportunity cost of both portfolios. We therefore focus only on the problem of GMV portfolio, but our method can be easily extended to MV portfolio.

The error of sample covariance may have large effect on the GMV portfolio performance, particulary so in situations with a large number of assets. As exemplified in Fan *et al.*. (2009)), for 2000 candidate assets, the covariance matrix involves over 2,000,000 parameters needing to be estimated, even each element in the covariance matrix is estimated with the accuracy of order $O(T^{-0.5})$ which is 0.05 if sample size $T = 400$ (1.5 years' daily data), the aggregated error could be very large resulting in devastating effect in portfolio selection. Three types of approaches have been proposed to solve the problem. The first is to use a factor structure for the asset returns to reduce the number of free parameters. Sharpe (1963) estimated the covariance matrix by using a single-factor market model, which reduces the number of parameters to be estimated dramatically. However, a single factor may not capture all of the covariations among the assets, resulting the estimated covariance matrix systematically biased. Multi-factor models were applied to overcome this drawback, but there is no consensus on the number and specification of the factors. It could be based on the economic theory, as implied by CAPM (Sharpe (1963)), ICAPM (Merton (1980)), or based on empirical work, like macroeconomic factors, industry factors, firm-characteristic-based factors and combined thereof (Chen *et al.*. (1986); Fama and French (1993)), or based on statistical procedure, like factor analysis or principal components analysis (Connor and Korajczyk (1998)). Chan *et al.* (1999) compared performance of different factor models in portfolio optimization problem and found that there was no clear favorite model specifications. The second method for improving the estimation of covariances matrix is to use the convex combinations of the classical sample estimator and a shrinkage target. The shrinkage targets could be an indentity matrix, the covariance matrix corresponding to a single or multi-factor model, or a covariance matrix with equal correlations (for example, see Frost and Savarino (1986); Ledoit and Wolf (2003)). Motivated by the fact that GMV optimal portfolio weights depend on the inverse covariance matrix, Kourtis, Dotsis and Markellos (2012) shrank the inverse covariance matrix directly instead of the covariance matrix. They proposed a linear combination of the traditional estimator of the inverse covariance matrix and a target matrix. The targets include the identity, the inverse covariance matrix generated by the 1-factor model of Sharpe (1963) and a weighted sum of these two matrices. Besides the above two types of methods, the third method to reduce the estimation error risk of parameters on the portfolio optimization is to impose constraints on the portfolio weights (for example, see Jagannathan and Ma (2003); Demiguel *et al.* (2009a)). Jagannathan and Ma (2003) proved that certain constraints on the portfolio weights was equivalent to a form of shrinkage estimation of covariances matrix.

Too little structure assumption on the covariation among the asset returns leads to poor performance of the estimator in small samples. Imposing some structure on the estimator is a natural cure. Particular forms of the structure in the above shrinkage estimation are based on financial theory, which have economic interpretations and are very familiar to financial academia, therefore they are easier to be accepted. However, there is no consensus on the identity and the number of factors, choosing a factor model is an art. Here we propose to shrink the off-diagonal elements of the inverse covariance matrix to zeros. It can be proved that under the assumption of multivariate normally distributed among the asset returns, the zeros of the inverse covariance imply that the corresponding asset returns are independent given the other asset returns, which is called conditional independence (Dawid (1980)). Besides simplifying the structure of the covariation among asset returns, conditional independence properties can be inspected visually by probabilistic graphical models. The asset returns is represented by nodes in the graph model, some of them are connected by links. The graph provides a visual way of understanding the joint distribution of the asset returns. In graph models, the absence of a link between two nodes means that two corresponding returns are conditionally independent given the other asset returns in the graph. It makes sense to constrain some returns to be of conditional independence which play an important role in pattern recognition and machine learning. Simulations and empirical data analysis show that sparse inverse covariance constraints reduce the estimation error of the covariance matrix, leading the resultant portfolio often have relatively better out-of-sample performance in terms of Sharpe ratios, variances and turn-overs than strategies with existing shrinkage covariance and inverse covariance estimates ((Ledoit and Wolf 2003, 2004), Kourtis, Dotsis and Markellos (2012)).

The rest of this article is organized as follows. In section 2, we propose a new GMV portfolio strategy by shrinking the off-diagonal elements of the inverse covariance matrix to zeros in the estimation of covariance matrix. In section 3, we prove some properties of the proposed portfolio strategies. In section 4, we compare the out-of-sample performance of the new strategy to some existing methods by simulation study and empirical data analysis.

Section 5 is the conclusion. Proofs of the propositions and details of the algorithm are available in Appendices.

2. Method

2.1. Imposing Constraints

Let $w = (w_1, ..., w_N)^\top$ be the vector of portfolio weights, Σ be the covariance matrix of the asset returns. Then, the Global Minimum Variance (GMV) portfolio strategy can be formulated as follows,

$$\min_{w} w^\top \Sigma w\ , \tag{1}$$

$$\text{s.t. } w^\top 1 = 1, \tag{2}$$

In order to control the estimation risk of the covariance matrix which play an important role in the GMV framework, we constrain the L_1 norm of the inverse covariance matrix in the estimation of the covariance matrix. Specifically, let $\Theta = \Sigma^{-1}$, S be the empirical covariance matrix, the problem is to maximize the penalized log-likelihood

$$\log \det\Theta - tr(S\ \Theta) - \rho\|\Theta\|_1 \tag{3}$$

over nonnegative definite matrices Θ. Here, tr denotes the trace, $\|\Theta\|_1$ is the L_1 norm or the sum of the absolute values of the elements of Σ^{-1} and ρ is the penalty coefficient. With the increment of ρ, we will get sparse inverse covariance matrix estimator. It can be proved that under some assumptions the zeros in the inverse covariance matrix imply that the corresponding two returns are conditionally independent. To be more specific, we assume that the asset returns have a multivariate Gaussian distribution with mean μ and covariance matrix Σ. The Gaussian distribution has a property that the inverse covariance matrix Σ^{-1} contains information about the partial covariances which is the covariances between two returns given the other returns. Specifically, if the ij^{th} component of Σ^{-1} is zero, then the i^{th} and the j^{th} asset return are conditionally independent given the other returns. So it makes sense to impose an L_1 penalty on Σ^{-1} to increase its sparsity.

In addition, the conditional independence can be visually checked by undirected probabilistic graphical models. In (undirected) graphical models, each node or vertex represents the future return of an asset, and each edge joining some pairs of the nodes (vertices) indicates the correlations among the asset returns. The absence of an edge between two nodes means that the corresponding two returns are conditionally independent given the other asset returns. Figure 1 gives an example of a graphical model for 10 assets. The absence of the edge between node 1 and node 4 denotes that

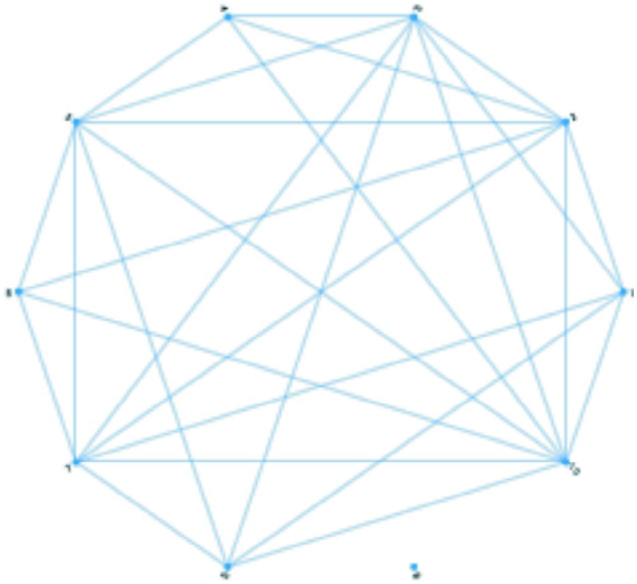

Figure 1: Undirected Graph Model for 10 Assets

returns for asset 1 and asset 4 are conditionally independent given other asset returns. And the shorthand notation for conditionally independence is as follows, $X_1 \perp X_4|\text{rest}$. Asset 9 is isolated, meaning asset 9 has nothing to do with other assets given other asset returns.

Friedman *et al.* (2007) developed an algorithm which is called graphical lasso to solve the convex optimization problem (3), the details of the algorithm can be found in Appendix A.

2.2. *Structure Calibration*

Graphical model gives a visual way of understanding the joint distribution of the entire set of asset returns. Sparse graphs are preferred for their simplified structures and resultant easy interpretations. In order to determine how sparse or which edge should be omitted from the graph, we can change the values of the penalty coefficient *Á* in expression (3) based on cross validation method.

To be more specific, let τ be the estimation window width, it usually sets to be 60 months, 120 months, or 180 months sample returns. We delete the i^{th} $(1 \leq i \leq \tau)$ observation from the estimation sample first, then calculate the covariance matrix estimate based on the $\tau - 1$ sample and a given ρ according to equation (3), then substitute $\hat{\Sigma}$ into equation (1) and equation

(2) to get portfolio weights, finally calculate the out-of-sample mean (variance, Sharpe ratio) by using the deleted i^{th} observation. As did in DeMiguel (2009), the out-of-sample mean, variance and Sharpe ratio are defined as follows:

$$\hat{\mu} = \frac{1}{\tau}\sum_{i=1}^{\tau} \hat{w}_{(i)}^{\top} r_i, \tag{4}$$

$$\hat{\sigma}^2 = \frac{1}{\tau - 1}\sum_{i=1}^{\tau} (\hat{w}_{(i)}^{\top} r_i - \hat{\mu})^2, \tag{5}$$

$$\widehat{SR} = \frac{\hat{\mu}}{\hat{\sigma}} \tag{6}$$

where $\hat{w}_{(i)}$ is calculated based on $\tau - 1$ sample returns without the i^{th} sample. Note that these quantities are based on the tuning parameter ρ, we chose ρ to get the best out-of-sample performance, that is to say, maximum out-of-sample mean or sharp ratio, or minimum out-of-sample variance.

3. Properties

We will show in this section the properties of new portfolio strategy based on the shrinkage inverse covariance matrix estimator.

Proposition 1: In our proposed framework consists of expression (1), (2) and (3), Markowitz strategy of GMV portfolio is obtained by setting $\rho = 0$, while adjusted $1/N$ strategy is reached by $\rho = \infty$.

This proposition shows that two popular strategies can be derived from our framework by changing the values of penalty coefficient. If $\rho = 0$, then there is no constraint on the sparsity of the inverse covariance, the corresponding graph model is fully edged which means that any two pairs of the nodes are connected. Therefore, the solution to (3) is the classical maximum likelihood estimate S. Plugging S into (2), we obtain the classical Markowitz strategy. While if $\rho = \infty$, then the constraint forces $\hat{\Theta}$ to be diagonal, implying the asset returns are mutually independent. The corresponding graph model consists of isolated nodes without any links. The estimate of covariance S from (3) is diagonal, resulting the strategy based on diagonal S is similar to $1/N$ strategy but adjusted according to the estimates of variances of the asset returns.

Proposition 2: If N asset returns are multivariate normally distributed, then the returns r_i and r_j are conditionally independent if and only if the ij^{th} element in the inverse covariance matrix Σ is zero.

This proposition explains the relationship between conditional independence and zeros in the inverse covariance matrix Σ^{-1}. For normal random variables, their third and higher-order joint cumulants which are defined as the natural logarithm of the moment-generating function are identically zero, therefore conditional dependence among them is expressed through their partial correlations, calculated from Σ^{-1}. So r_i and r_j are independent conditional on all the other returns if and only if their partial correlation is zero. For details of calculating partial correlations from Σ^{-1} can be found in appendix B.

Proposition 3: There exists a local optimizer $\hat{\Theta}$ for the penalized loglikelihood (3) optimization problem with a certain rate of convergence. If $\theta_{ij} = 0$, then $Pr(\hat{\theta}_{ij} = 0) \to 1$, and other elements of $\hat{\Theta}$ have the same limiting distribution as the maximum likelihood estimator on the true covariance matrix.

Proposition 3 was proved under some mild conditions by Lam and Fan (2009), indicating that the constrained estimator of covariance selects the right graph with probability tending to one and at the same time gives a root-n consistent estimator of the precision matrix.

4. Out-of-Sample Evaluation

4.1. Benchmarks and Criterion for Comparison

In this paper, we chose 7 popular strategies as benchmarks and compare the performance of these strategies with the proposed strategy in this paper. $1/N$ strategy allocates all the money equally among different assets. MINU and MINC are strategies without and with short-sales constraints respectively. LW_{id} strategy is based on shrinkage covariance estimates which are combinations of the empirical covariance estimator and the identity matrix (Ledoit and Wolf (2004)). LW_{if} strategy is similar to LW_{id}, but applies the covariance matrix corresponding to a one-factor model for the asset returns as the shrinkage target (Ledoit and Wolf (2003)). KDM_{id} strategy is based on shrinkage inverse covariance estimates which are combinations of the empirical inverse covariance estimator and the identity matrix (Kourtis, Dotsis and Markellos (2012)). KDM_{if} strategy is similar to KDM_{id}, but chooses the inverse covariance matrix estimator resulting from the one-factor model for the asset returns as the shrinkage target (Kourtis, Dotsis and Markellos (2012)). *SICP* strategy proposed in this paper represents portfolios selected based on sparse inverse covariance matrix estimates.

We used "rolling widow" method to calculate the out-of sample performance for the above different strategies. The window width was chosen to be 120 month, which is 10 years. we tested 180 months and 60 months also to show the results are robust. Let τ denote the window width, T to be the overall observations. For $t = \tau, \tau + 1, \ldots, T - 1$, we use the data points in the window $[t - \tau + 1, t]$ to calculate the optimal portfolio weights $\hat{w}_t$ and estimate the out-of-sample performance as follows.

$$\hat{\mu} = \frac{1}{T - \tau} \sum_{t=\tau}^{T-1} \hat{w}_{(i)}^{\top} r_{t+1}, \tag{4}$$

$$\hat{\sigma}^2 = \frac{1}{T - \tau - 1} \sum_{t=\tau}^{T-1} (\hat{w}_{(i)}^{\top} r_{t+1} - \hat{\mu})^2, \tag{5}$$

$$\widehat{SR} = \frac{\hat{\mu}}{\hat{\sigma}} \tag{9}$$

we also estimate the out-of-sample turnover which is defined as

$$\widehat{TO} = \frac{1}{T - \tau - 1} \sum_{t=\tau}^{T-1} \| \hat{w}_{(i)}^{\top} r_{t+1} - \hat{\mu} \|)^2,$$

where $\hat{w}_{t+1}$ is the desired portfolio weight at $t + 1$ after rebalancing, and $\hat{w}_{t^+}$ is the portfolio weight before rebalancing at $t + 1$.

4.2. Real Data Analysis

Descriptions of the four real data sets are given in Table 1.

Table 2 presents the out-of-sample Sharpe ratios, variances and turnovers for the four data sets 10Ind, 30Ind, 100FF, CRSP under different portfolio strategies. We can see that SICP has the largest out-of-sample Sharpe ratio for data 10Ind, 30Ind, 100FF, and the second largest for data CRSP. For data 10Ind, 30Ind, 100FF, other shrinkage methods have similar Sharpe ratios and generally smaller than those of SICP, though the differences are not always statistically significant. For data CRSP, SICP has

Table 1: The data sets of monthly asset returns analyzed in this study

No.	*Data Set*	*Abbreviation*	*N*	*Time period*	*Source*
1.	Ten industry portfolios	10Ind	10	01/1970-12/2013	K. French
2.	Thirty industry portfolios	30Ind	30	01/1970-12/2013	K. French
3.	One hundred Portfolios	100FF	100	01/1970-12/2013	K. French
4.	One hundred Random Portfolio	CRSP	100	12/1971-12/2013	CRSP

a This table lists four data sets of monthly asset returns. The first and second sets are industry portfolios. The third set is formed on size and book-to-market. The forth set selects 100 stocks randomly from CRSP (The Center for Research in Security Prices).

b *Source:* http : //mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Table 2: Out-of-sample performance measures for real data sets

	Sharpe ratio				*Variance* ($\times 10^{-3}$)				*Turnover*			
Source	*10Ind*	*30Ind*	*100FF*	*CRSP*	*10Ind*	*30Ind*	*100FF*	*CRSP*	*10Ind*	*30Ind*	*100Ind*	*CRSP*
1/N	0.266	0.242	0.223	0.156	1.779	2.207	2.510	2.256	0.024	0.029	0.026	0.055
	[0.127]	[0.043]	[0.004]	[0.910]	[0.000]	[0.000]	[0.000]	[0.000]	[–]	[–]	[–]	[–]
MINU	0.324	0.268	0.154	0.082	1.247	1.374	6.229	11.042	0.155	0.464	7.604	5.525
	[0.614]	[0.117]	[0.008]	[0.324]	[0.198]	[0.067]	[0.000]	[0.000]	[–]	[–]	[–]	[–]
MINC	0.306	0.282	0.255	0.164	1.256	1.276	1.868	1.162	0.051	0.068	0.117	0.142
	[0.128]	[0.053]	[0.002]	[0.930]	[0.214]	[0.103]	[0.000]	[0.414]	[–]	[–]	[–]	[–]
LW_{id}	0.324	0.292	0.417	0.104	1.247	1.224	1.439	1.296	0.155	0.237	0.406	0.394
	[0.611]	[0.252]	[0.235]	[0.020]	[0.187]	[0.616]	[0.077]	[0.020]	[–]	[–]	[–]	[–]
LW_{if}	0.328	0.312	0.359	0.128	1.217	1.181	1.342	1.118	0.118	0.179	0.801	0.222
	[0.831]	[0.756]	[0.342]	[0.074]	[0.763]	[0.847]	[0.068]	[0.765]	[–]	[–]	[–]	[–]
KDM_{id}	0.324	0.291	0.299	0.083	1.200	1.452	2.881	7.398	0.155	0.209	0.267	4.301
	[0.094]	[0.358]	[0.250]	[0.303]	[0.000]	[0.013]	[0.062]	[0.000]	[–]	[–]	[–]	[–]
KDM_{if}	0.328	0.307	0.247	0.096	1.244	1.273	2.764	1.522	0.128	0.209	3.695	0.144
	[0.578]	[0.666]	[0.019]	[0.083]	[0.004]	[0.286]	[0.000]	[0.004]	[–]	[–]	[–]	[–]
SICP	0.330	0.318	0.373	0.161	1.211	1.192	1.275	1.101	0.099	0.114	0.726	0.234
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[–]	[–]	[–]	[–]

a The numbers in square brackets are *p*-values of the portfolio Sharpe ratios and variances for a strategy is different from that for SICP strategy. The p-values are computed using the stationary bootstrap method proposed by Ledoit *et al.* (2008).

larger Sharpe ratio than other shrinkage strategies and most of the time are statistically significant. As for out-of-sample variance, SICP has the smallest variance for data 30Ind, 100FF, CRSP and the second smallest for data 10Ind, the differences are often statistically significant. In addition, the turnover values for SICP are often smaller compared other shrinkage strategies for 10Ind and 30Ind. However, for data 100Ind and CRSP, the turnover values for SICP are bigger than shrinkage strategy which shrink covariance or inverse covariance to the identity matrix, but smaller than that of the strategy which shrink the covariance and inverse covariance based on the market factor model.

The graph structure of applying SICP to data 10Ind is shown in Figure 2, for 15 different values of the penalty parameter ρ in equation (3). We can see that the graph structure is becoming more sparser with the decrement of $L1$ norm of the inverse covariance matrix. The extent of sparsity is determined according to the out-of-sample performance of the portfolio described in section 2.2.

4.3. Simulation Study

This section investigates the out-of-sample performance by simulation study. We generate data based on factor models. Let r_i be the return rate of asset i, r_f be the risk-free return rate, f be factors, and B be the factor loading matrix. The factor model can be represented as follows,

$$r = 1r_f + Bf + \epsilon. \tag{10}$$

We use one factor $f = (R_m - r_f)^\top$, three factors $f = (R_m - r_f, SMB, HML)^\top$ and five factors $f = (R_m - r_f, SMB, HML, RMW, CMA)^\top$ respectively. Where R_m is the return of the whole stock market, $R_m - r_f$ is the first factor which is the same as CAPM (capital asset pricing model), SMB is the difference between the returns on small stocks and big stocks, HML is the difference between returns of high and low B/M stocks, the above three factors are the same as Fama and French (1993) 3-factor model. RMW is the difference between the returns on diversified portfolios of the stocks with robust and weak profitability, and CMA is the difference between the returns on diversified portfolios of the stocks of low/conservative and high/aggressive investment firms. Those two factors are the same with Fama and French (2015) 5-factor model. In one factor model, the factor loading is normally distributed with $\mu_b = 1.077$, $\sigma_b = 0.034$, and the factor is normally distributed with $\mu_f = (8/12)\%$ and $\mu_f = (16/\sqrt{(12)})\%$ which is similar as MacKinlay and Pástor (2000). In 3-factor model, the factor loadings are normally distributed with $\mu_b = (1.085, 0.489, 0.365)$ and

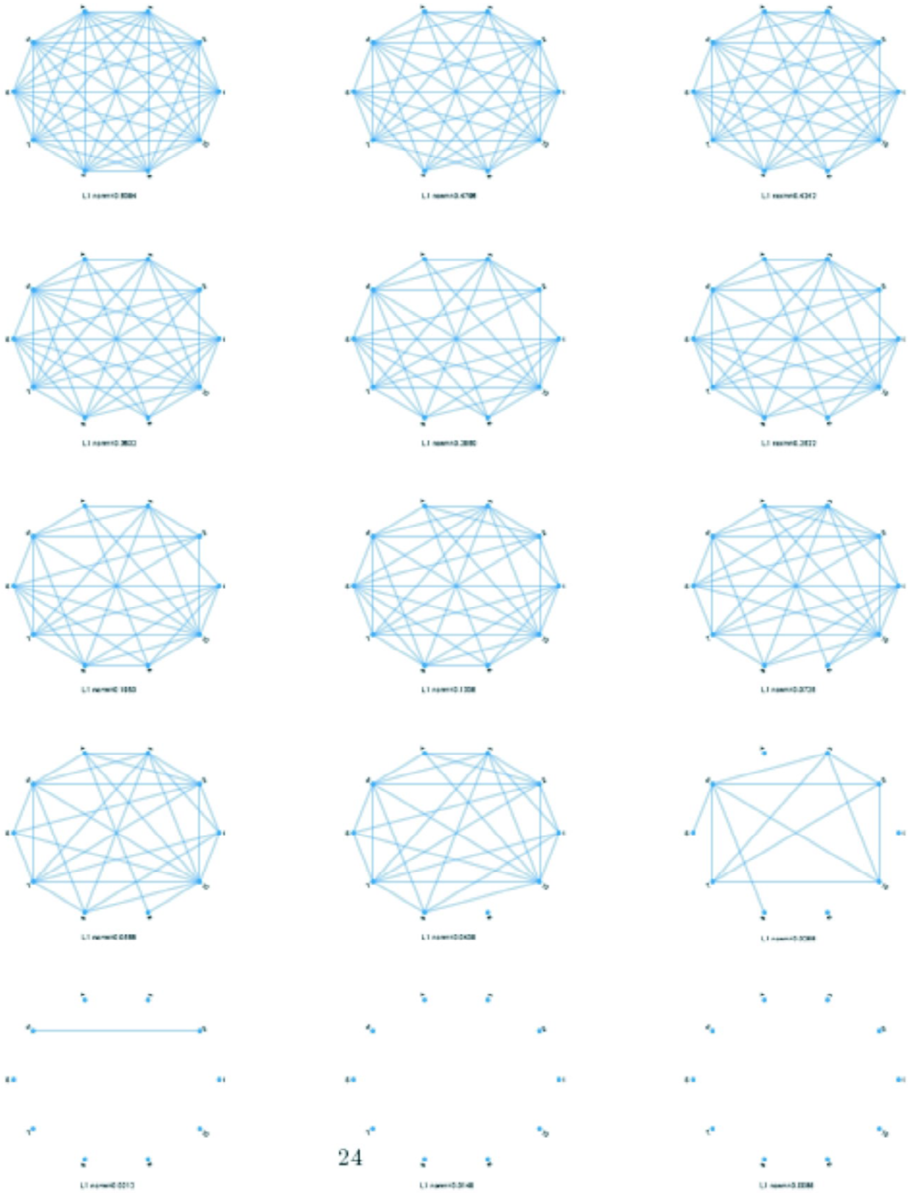

Figure 2: Solution paths of several norm constrained portfolios.

$$\Sigma_b = \begin{pmatrix} -0.010 & -0.001 & 0.001 \\ -0.001 & 0.250 & 0.011 \\ 0.001 & 0.011 & 0.204 \end{pmatrix}$$

The 3 factors are normally distributed with $\mu_f = (0.488, 0.195, 0.484)$ and

$$\Sigma_f = \begin{pmatrix} -0.010 & -0.001 & 0.001 \\ -0.001 & 0.250 & 0.011 \\ 0.001 & 0.011 & 0.204 \end{pmatrix}$$

In 5-factor model, the factor loadings are normally distributed with μ_b = (1.073, 0.521, 0.343, 0.026, –0.078) and

$$\Sigma_b = \begin{pmatrix} 0.011 & 0.008 & -0.001 & 0.015 & 0.016 \\ 0.008 & 0.260 & -0.049 & 0.046 & 0.096 \\ -0.001 & -0.049 & 0.163 & 0.016 & -0.011 \\ 0.015 & 0.046 & 0.016 & 0.125 & 0.115 \\ 0.016 & 0.096 & -0.011 & 0.115 & 0.167 \end{pmatrix}$$

The 5 factors are normally distributed with μ_f = (0.495, 0.230, 0.484, 0.294, 0.400) and

$$\Sigma_f = \begin{pmatrix} 20.819 & 3.179 & -6.322 & -1.975 & -4.394 \\ 3.179 & 10.317 & -1.741 & -2.959 & -0.465 \\ -6.321 & -1.741 & 9.450 & 1.470 & 4.692 \\ -1.975 & -2.959 & 1.470 & 5.597 & -0.097 \\ -4.394 & -0.465 & 4.692 & -0.097 & 4.405 \end{pmatrix}$$

We generate T = 10120 monthly returns for N = 10, 30, 100 assets based on one, three and five factor models respectively. The out-of-sample performance are measured by rolling window method and window width τ is set to be 120 months, which are the same as empirical data analysis. The out-of-sample Sharpe ratios, Variances and Turnovers based on different simulated data sets are presented on table 3, table 4 and table 5. We can see that SICP has the largest out-of-sample Sharpe ratio and relative lower out-of-sample variance across all data sets for different factor models, the differences are generally statistically significant. Furthermore, with the increment of the number of factors and the number of assets, the advantages of SICP strategy become more and more obvious. Particularly, table 5 shows that compared with other strategies, SICP strategy increases out-of-sample Sharpe ratios and decreases out-of-sample variances dramatically and statistical significantly for the data set N = 100 generated from five factor model.

Table 3: Out-of-sample performance measures for simulated data sets based on onefactor model

	Sharpe ratio			*Variance* ($\times 10^{-3}$)			*Turnover*		
Size	*10*	*30*	*100*	*10*	*30*	*100*	*10*	*30*	*100*
1/N	0.102	0.099	0.110	2.457	2.406	2.325	0.044	0.046	0.046
	[0.352]	[0.562]	[0.187]	[0.000]	[0.000]	[0.000]	[–]	[–]	[–]
MINU	0.066	0.034	0.013	2.028	1.823	4.699	0.097	0.264	2.509
	[0.021]	[0.025]	[0.174]	[0.065]	[0.000]	[0.000]	[–]	[–]	[–]
MINC	0.086	0.071	0.085	2.101	1.835	1.633	0.064	0.088	0.125
	[0.268]	[0.087]	[0.088]	[0.000]	[0.000]	[0.000]	[–]	[–]	[–]
LW_{id}	0.102	0.099	0.110	2.457	2.406	2.324	0.091	0.233	0.046
	[0.366]	[0.559]	[0.185]	[0.209]	[0.002]	[0.000]	[–]	[–]	[–]
LW_{if}	0.066	0.050	0.033	2.019	1.611	0.899	0.090	0.167	0.258
	[0.013]	[0.011]	[0.001]	[0.119]	[0.549]	[0.000]	[–]	[–]	[–]
KDM_{id}	0.101	0.099	0.110	2.457	2.399	2.312	0.097	0.264	0.198
	[0.361]	[0.526]	[0.022]	[0.000]	[0.500]	[0.000]	[–]	[–]	[–]
KDM_{if}	0.102	0.050	0.032	2.457	1.613	0.899	0.091	0.168	0.258
	[0.320]	[0.016]	[0.002]	[0.090]	[0.050]	[0.000]	[–]	[–]	[–]
SICP	0.101	0.099	0.110	2.438	2.392	2.280	0.072	0.143	0.203
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[–]	[–]	[–]

a The numbers in square brackets are *p*-values of the portfolio Sharpe ratios and variances for a strategy is different from that for SICP strategy. The *p*-values are computed using the stationary bootstrap method proposed by Ledoit and Wolf (2008).

4.4. Conclusion

Many approaches had been proposed to solve the problem of estimation risk in GMV portfolio optimization. Among them, shrinking the covariance matrix or inverse covariance matrix to a particular target has been used. The targets are selected according to the financial theory and there is no consensus on the target matrix. In this paper, we propose to shrink the inverse covariance matrix to a sparse matrix. The sparsity structure of the inverse covariance matrix are determined by out-of-sample performances. Specifically, we use the tuning parameter ρ in the penalized likelihood to adjust the sparsity structure of the inverse covariance matrix estimates, the criterion for choosing the value of ρ is the out-of-sample Sharpe ratios, variances and turnovers of the portfolios. The main advantage of our new method is that it is totally data driven. Furthermore, it has been approved that the zeros in the inverse covariance matrix imply that the corresponding asset returns are independent given the other asset returns if the asset returns are normally distributed. Therefore it makes sense to impose sparsity constraints to the inverse covariance matrix. With different values of the

Table 4: Out-of-sample performance measures for simulated data sets based on threefactor model

	Sharpe ratio			*Variance* ($\times 10^{-3}$)			*Turnover*		
Size	*10*	*30*	*100*	*10*	*30*	*100*	*10*	*30*	*100*
1/N	0.119	0.097	0.136	1.675	2.671	2.868	0.038	0.037	0.048
	[0.518]	[0.464]	[0.007]	[0.000]	[0.000]	[0.000]	[–]	[–]	[–]
MINU	0.115	0.056	0.078	1.434	0.304	6.262	0.153	0.639	3.082
	[0.444]	[0.075]	[0.003]	[0.190]	[0.000]	[0.000]	[–]	[–]	[–]
MINC	0.119	0.108	0.003	1.627	1.927	1.978	0.05	0.088	0.134
	[0.969]	[0.661]	[0.003]	[0.000]	[0.000]	[0.000]	[–]	[–]	[–]
LW_{id}	0.119	0.105	0.136	1.449	1.154	2.868	0.09	0.251	0.048
	[0.985]	[0.064]	[0.007]	[0.002]	[0.000]	[0.000]	[–]	[–]	[–]
LW_{if}	0.115	0.105	0.136	1.430	0.537	1.373	0.130	0.319	0.321
	[0.071]	[0.096]	[0.005]	[0.038]	[0.000]	[0.000]	[–]	[–]	[–]
KDM_{id}	0.115	0.100	0.141	1.434	1.976	2.678	0.153	0.087	0.189
	[0.455]	[0.497]	[0.013]	[0.189]	[0.000]	[0.000]	[–]	[–]	[–]
KDM_{if}	0.110	0.106	0.105	1.438	1.332	1.966	0.154	0.391	0.820
	[0.038]	[0.048]	[0.022]	[0.000]	[0.000]	[0.000]	[–]	[–]	[–]
SICP	0.119	0.115	0.188	1.420	1.067	1.715	0.107	0.228	0.167
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[–]	[–]	[–]

a The numbers in square brackets are *p*-values of the portfolio Sharpe ratios and variances for a strategy is different from that for SICP strategy. The *p*-values are computed using the stationary bootstrap method proposed by Ledoit and Wolf (2008).

Table 5: Out-of-sample performance measures for simulated data sets based on fivefactor model

	Sharpe ratio			*Variance* ($\times 10^{-3}$)			*Turnover*		
Size	*10*	*30*	*100*	*10*	*30*	*100*	*10*	*30*	*100*
1/N	0.063	0.095	0.136	2.736	2.599	2.851	0.045	0.047	0.048
	[0.874]	[0.418]	[0.096]	[0.000]	[0.000]	[0.000]	[–]	[–]	[–]
MINU	0.064	0.093	0.053	1.855	1.814	6.804	0.104	0.308	3.171
	[0.702]	[0.341]	[0.005]	[0.751]	[0.482]	[0.000]	[–]	[–]	[–]
MINC	0.056	0.094	0.124	2.031	2.177	2.087	0.044	0.092	0.139
	[0.38]	[0.223]	[0.003]	[0.000]	[0.000]	[0.064]	[–]	[–]	[–]
LW_{id}	0.065	0.095	0.136	1.852	1.952	2.851	0.099	0.249	0.048
	[0.76]	[0.059]	[0.090]	[0.473]	[0.000]	[0.000]	[–]	[–]	[–]
LW_{if}	0.062	0.103	0.129	1.821	1.57	1.397	0.089	0.198	0.334
	[0.479]	[0.665]	[0.297]	[0.054]	[0.000]	[0.000]	[–]	[–]	[–]
KDM_{id}	0.064	0.096	0.139	1.855	2.529	2.687	0.104	0.046	0.190
	[0.695]	[0.463]	[0.143]	[0.768]	[0.000]	[0.000]	[–]	[–]	[–]
KDM_{if}	0.064	0.094	0.095	1.857	1.607	2.099	0.104	0.211	0.899
	[0.669]	[0.338]	[0.063]	[0.849]	[0.000]	[0.437]	[–]	[–]	[–]
SICP	0.066	0.109	0.155	1.861	1.853	2.009	0.088	0.170	0.104
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[–]	[–]	[–]

a The numbers in square brackets are *p*-values of the portfolio Sharpe ratios and variances for a strategy is different from that for SICP strategy. The *p*-values are computed using the stationary bootstrap method proposed by Ledoit and Wolf (2008).

tuning parameter ρ, different portfolio optimization strategies are derived, for example, naive strategy $1/N$ and classical plug-in strategy are shown to be the special cases of our new method. Last but not the least, empirical data analysis and simulation studies show that new strategies usually have better out-of-sample performances in terms of Sharpe ratios, variances and turnovers, and often statistically significant, especially when the number of the assets is large.

5. Acknowledgments

This work is partially supported by China NSF grant and Shanghai Pujiang Talent Program.

References

Brandt, M.W., 2007. Portfolio Choice Problems. In: Ait-Sahalia,Y., Hansen, L.P. (Eds.). *Handbook of Financial Econometrics*. Vol. 1. Elsevier, Amsterdam.

Chan, L., Karceski, J., Lakonishok, J., 1999, On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies* 12, 937-974.

Chen, N.F., Roll R., Ross S.A., 1986, Economic forces and the stock market. *Journal of Business* 59, 383-403.

Connor, G., Korajczyk, R., 1998. Risk and return in an equilibrium APT: Application of a new test methodology, *Journal of Financial Economics* 21, 255-290.

Dawid, A.P., 1980, Conditional independence for statistical operations. *Annals of Statistics* 8, 598-617.

Demiguel, V., Garlappi, L., Nogales, F.J.,Uppal, R., 2009a. Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. Management Science 22, 1915–1953.

Demiguel, V., Garlappi,L., Uppal, R., 2009b. Optimal versus naive diversification: How inefficient is the 1/*N* portfolio strategy? *Review of Financial Studies* 55, 798–812.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.

Fan, J., Zhang, J., Yu, K., 2009. Asset allocation and risk assessment with gross exposure constraints for vast portfolios. *The Annals of Statistics* 25, 1425-1432.

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

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

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

Jobson, J.D., Korkie, B., 1980. Estimation of Markowitz efficient portfolios, *Journal of the American Statistical Association* 75, 544-554.

Kourtis, V., Dotsis G., Markellos R. N., 2012. Parameter uuncertainty in portfolio selection: Shrinking the inverse covariance matrix. *Journal of Banking and Finance* 36, 2522-2531.

Lam, C., Fan, J.Q., 2009. Sparsistency and rates of convergence in large covariance matrix estimation. *The annals of Statistics* 37, 4254-4278.

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

Ledoit, O., Wolf, M., 2004. Honey, I shrunk the sample covariance matrix: problems in mean-variance optimization. *Journal of Portfolio Management* 30, 110-119.

Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance* 15, 850–859.

MacKinlay, A.C., Pástor, L., 2000. Asset pricing models: implications for expected returns and portfolio selection. *Review of Financial Studies* 13, 883-916.

Markowitz, H., 1952. Portfolio selection. The journal of Finance 7, 77–91.

Merton, R. C., 1980. On estimation the expected return on the market. *Journal Financial Economics* 8, 323-361.

Michaud, R.O., 1989. The Markowitz optimization enigma: Is optimized optimal? *Financial Analyst Journal* 45, 31-42.

Sharpe, W., 1963. A simplified model for portfolio analysis. *Management Science* 9, 277-293.

Appendix A: The Algorithm

Friedman *et al.* (2007) developed an algorithm which is called graphical lasso to solve the convex optimization problem

$$\log \det\Theta - tr(S\,\Theta) - \rho\|\Theta\|_1 \tag{A1}$$

where $\Theta = \Sigma^{-1}$, S is the empirical covariance matrix, tr denotes the trace, $\|\Theta\|_1$ is the L_1 norm or the sum of the absolute values of the elements of Σ^{-1} and ρ is the penalty coefficient.

Let W be the estimate of Σ, partitioning W and S as following:

$$W = \begin{bmatrix} W_{11} & \omega_{12} \\ \omega_{12}^T & \omega_{12} \end{bmatrix} \quad S = \begin{bmatrix} S_{11} & s_{12} \\ s_{12}^T & s_{12} \end{bmatrix}$$

The details of graphical lasso algorithm are as following:

1. Initialize $W = S + \rho \cdot I$, I is the identity matrix. The diagonal of W remains unchanged in what follows.
2. Repeat for $j = 1, 2, \ldots p, 1, 2, \ldots p, .$ until convergence:
   - Partition the matrix$W$ into part 1: all but the j th row and column, and part 2: the j th row and column.
   - Solve the estimating equations $W_{11}\beta - s_{12} + \rho \cdot \text{Sign}(\beta) = 0$ using the cyclical coordinate-decent algorithm (A2) for deriving the estimate of β. Specifically, let $V = W_{11}$ update $\hat{\beta}_j$ by

$$\hat{\beta}_j \leftarrow S\left(s_{12j} - \sum_{k \neq j} V_{kj}\hat{\beta}_k, \rho\right) \Big/ V_{jj} \tag{A2}$$

     for $j = 1, 2, \ldots, p-1, \ldots$, where S is the soft-threshold operator:

$$S(x, t) = \text{sign}(x)(|x| - t)_+$$

   - Update $\omega_{12} = W_{11}\hat{\beta}$.
3. In the final cycle (for each j) solve for $\theta_{12} = -\hat{\beta} \cdot \hat{\theta}_{22}$, with $1/\hat{\theta}_{22} = \omega_{22} - \omega_{12}^T\hat{\beta}$.

   The graphical lasso algorithm is very fast, and can solve a moderately sparse problem with 1000 nodes in less than a minute.

Appendix B: Proof of proposition 2

Assuming that the asset returns are multivariate normally distributed, we prove that if the ij th component of $\Theta = \Sigma^{-1}$ is zero, then the i th asset and the

j^{th} asset are conditionally independent, given the other asset returns. Let $R = (r_1, r_2, \ldots r_N)$, we partition $R = (R_S, R_{-S})$, where $R_S = (r_i, r_j)$, $\forall 1 \leq i, j \leq N$, R_{-S} contains the remaining asset returns. The partial covariance between asset i and asset j conditional on the remaining asset returns is

$$\Sigma_{ij|-S} = \Sigma_{SS} - \Sigma_{S,-S}\Sigma_{-S,-S}^{-1}\Sigma_{-S,S}$$

The partial correlation coefficient $\rho_{ij|-S}$ can be computed from the above partial covariance.

$$\rho_{ij|-S} = \frac{\sigma_{ij|-S}}{(\sigma_{ij|-S}\sigma_{jj|-S})^{1/2}}$$

$$= \frac{\sigma_{ij} - \Sigma_{i,-S}\Sigma_{-S,-S}^{-1}\Sigma_{-Sj}}{\{(\sigma_{ii} - \Sigma_{i,-S}\Sigma_{-S,-S}\Sigma_{-S,i})(\sigma_{ij} - \Sigma_{j,-S}\Sigma_{-S,-S}\Sigma_{-S,j})\}^{1/2}} \quad \text{(A3)}$$

Based on Cramer's rule in linear algebra, the (i, j) element of Σ^{-1} is $(-1)^{i+j}\Sigma_{ij}/|\Sigma|$, where Σ_{ij} defined as the (i, j) minor of Σ is the determinant of the submatrix formed by deleting the i-th row and the j-th column. The "correlationized" version of Σ^{-1} is $(-1)^{i+j}\Sigma_{ij}/(\Sigma_{ii}\Sigma_{jj})^{1/2}$.

According to the formula

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = \left|A_{11} - A_{12}A_{22}^{-1}A_{21}\right| \cdot \left|A_{22}\right|$$

for the determinant of a partitioned matrix for which A_{22}^{-1} exists. By making the row and column interchanges that bring σ_{ii} to the (1, 1) position of $\Sigma_{-i,-i}$, we derive that

$$\Sigma_{ii} = (-1)^{2(i-1)}\begin{vmatrix} \sigma_{ii} & \Sigma_{i,-S} \\ \Sigma_{-S,i} & \Sigma_{-S,-S} \end{vmatrix}$$

$$= (\sigma_{ii} - \Sigma_{i,-S}\Sigma_{-S,-S}^{-1}\Sigma_{-S,i})|\Sigma_{-S,-S}|$$

similarly, we can get the expression for Σ_{jj}, while Σ_{ij} is as follows:

$$\Sigma_{ij} = (-1)^{i+j-1)}\begin{vmatrix} \sigma_{ij} & \Sigma_{i,-S} \\ \Sigma_{-S,j} & \Sigma_{-S,-S} \end{vmatrix}$$

$$= (-1)^{i+j-1}(\sigma_{i,j} - \Sigma_{i,-S}\Sigma_{-S,-S}^{-1}\Sigma_{-S,j})|\Sigma_{-S,-S}|$$

Substituting the expressions for Σ_{ii}, Σ_{jj} and Σ_{ij} into $(-1)^{i+j}\Sigma_{ij}/(\Sigma_{ii}\Sigma_{jj})^{1/2}$, we derive that the (i, j) element of the "correlationalized" version of Σ^{-1} equals $-\rho_{ij|S}$ as in equation (A3).