

Robust Portfolio Risk Minimization Using the Graphical Lasso

Tristan Millington & Mahesan Niranjana

Department of Electronics and Computer Science
University of Southampton
Highfield SO17 1BJ, Southampton,
United Kingdom
{T.Millington/M.Niranjana}@Southampton.ac.uk

Abstract. We apply the statistical technique of *graphical lasso* for inverse covariance estimation of asset price returns in Markowitz portfolio optimisation. Graphical lasso induces sparsity in the inverse covariance matrix, thereby capturing conditional independences between different assets. We show empirical results that not only the resulting minimum risk portfolio is robust, in that the variation in expected returns is reduced when a fraction of the data is assumed missing, but also enables the construction of a financial network in which groups of assets belonging to the same financial sector are linked.

Keywords: Portfolio Optimization, Graphical Lasso, Financial Network, Graphical Model, Covariance Estimation

1 Introduction

Portfolio optimisation and its variants have been of interest in empirical finance for decades, following the pioneering work of Markowitz [17]. This mean-variance optimisation problem defines a Pareto optimal frontier in the space of expected returns and the corresponding risks (variances) of a portfolio under the assumption that returns on the assets follow a multivariate Gaussian density. The formulation of such a model, under assumptions of no short selling, is given as follows:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \mathbf{w}^T \Sigma \mathbf{w} \\ & \text{subject to} && \sum_{i=1}^{i=N} w_i = 1 \\ & && w_i \geq 0 \\ & && \mathbf{w}^T \mathbf{m} \leq \rho \end{aligned} \tag{1}$$

where, \mathbf{w} represents the portfolio weights $[w_1, w_2 \dots w_N]^T$, \mathbf{m} and Σ the parameters of the Gaussian distributed asset returns, and ρ , the expected returns. Solving the above quadratic programming problem at different values of ρ yields the well known efficient frontier.

In practice, however, \mathbf{m} and Σ are not known and have to be estimated from data with the maximum likelihood estimates of

$$\widehat{\mathbf{m}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \quad (2)$$

$$\widehat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_t - \mathbf{m})(\mathbf{x}_t - \mathbf{m})^T \quad (3)$$

being the common choice. Here, T is the time window of data and $\mathbf{x}_t \in \mathcal{R}^N$ represents returns on the assets at time t .

In finance, these estimates from data are known to suffer robustness issues [18], [14], [4]. Financial data consists of occasional outliers to which maximum likelihood estimates are notoriously sensitive. Further, to estimate covariance matrices reliably we need a long enough window (T) of data, though due to non-stationarity in the markets, we may choose a small window. This (particularly when T and N are of similar values) can lead to the covariance matrix Σ being singular and non-invertible. The consequence of poor estimation of parameters is that the resulting portfolio can be unstable and produce poor out-of-sample performance, with extreme weights that are liable to have large changes over time [4]. In many cases these portfolios perform worse than a $1/N$ naive portfolio [8]. These issues are often addressed by regularisation, of which shrinkage estimation is a classic tool (*e.g.* [14]). Brodie *et al.* address this issue by regularisation using the l_1 (or *lasso*) penalty in an index tracking setting, deriving stable portfolios which are also sparse. Takeda *et al.* [19] use a combination of l_1 and l_2 regularisers to simultaneously induce sparsity and improve out-of-sample performance.

Estimation of covariance matrices with some desirable structure imposed on them falls under the field of structured matrix approximation (*e.g.* [7], [16]). In the financial domain, Fan *et al.* [10] involve 3-factor model in the estimation of a covariance matrix. They find the factor model improves the estimation of the precision matrix, but affects errors in the estimation of the covariance matrix less. In this setting, Friedman *et al.* [11], introduced the *graphical lasso* (glasso) as a way of inducing structure into covariance matrices. Specifically, they argued that a sparse inverse covariance matrix of Gaussian distributed data captures conditional independences between the variables. From this algorithm, a well-conditioned covariance matrix and a sparse precision matrix are produced even when T is approaching N . From this sparse precision matrix, a network can be extracted, with a zero in the matrix indicating a conditional independence and a non-zero value indicating a relationship. We make use of the well-conditioned covariance matrix and the precision matrix in this paper.

To start, we assume the asset returns follow a multivariate Gaussian distribution. A multivariate Gaussian distribution can be written as

$$f(\mathbf{x}, \mathbf{m}, \Sigma) = \frac{1}{(2\pi)^{\frac{T}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \Sigma^{-1}(\mathbf{x} - \mathbf{m})\right). \quad (4)$$

Suppose we have T measurements of N assets, denoted as $\mathbf{X} = (\mathbf{x}_1 \dots \mathbf{x}_N)^T$ where each \mathbf{x} is a T dimensional vector. The log-likelihood this set of measurements belongs to a Gaussian distribution with Σ and \mathbf{m} is

$$\log L(\mathbf{m}, \Sigma | \mathbf{X}) = -\frac{TN}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^T (\mathbf{x}_t - \mathbf{m})^T \Sigma^{-1} (\mathbf{x}_t - \mathbf{m}). \quad (5)$$

Substituting the maximum likelihood covariance estimators from (2) and (3) and discarding the constant gives

$$\log L(\mathbf{m}, \Sigma | \mathbf{X}) = \log |\Theta| - \text{tr}(\hat{\Sigma} \Theta) \quad (6)$$

where $\Theta = \Sigma^{-1}$.

Following Banerjee *et al.* [2], we add a L_1 penalty term to impose sparsity on the precision matrix:

$$\log \det \Theta - \text{tr}(\hat{\Sigma} \Theta) - \lambda \|\Theta\|_1 \quad (7)$$

with λ as a regularising parameter.

Friedman *et al.* [11] propose the glasso to maximise this function using block coordinate gradient descent. Pseudo-code for the glasso algorithm is shown in Figure 1.

We are not alone in looking for applications for the glasso. Goto *et al.* [12] use the glasso to construct a sparse precision matrix for portfolio hedging. Exploiting the sparsity allows for lower turnover in the portfolio hedging and gives a more predictable out-of-sample risk and return. Awoye [1] uses the glasso to estimate a covariance matrix for mean-variance portfolio optimisation, and compares its performance to existing covariance estimators with various constraints. Their portfolios designed using the glasso perform well when compared to other estimators, achieving a lower realised risk. We hope to exploit the lowered risk and increased robustness the glasso provides.

We are also not the first to construct networks on financial data. Mantegna [15] constructed a network from stock prices using a correlation matrix. Defining a distance metric using correlation coefficients, they construct a minimum spanning tree from the companies used to construct the Dow Jones Industrial Average. Companies in similar sectors are clustered in this minimum spanning tree. Boginski *et al.* [5] also use a correlation matrix to create a network. Using a threshold on the correlation coefficient to decide whether two companies are linked, they build a graph. They find the produced graph follows a power law rule when the threshold is set at a large enough value, and look at how varying the threshold changes the clusters in the graph. Huang *et al.* [13] follow a similar path, instead concentrating on the Chinese stock market. They use their network to classify the stocks and to test the stability of the market. We also wish to see how our constructed network can classify companies.

Input : Empirical Covariance Matrix $\hat{\Sigma}$
 Convergence Tuning Parameter t
 Regularisation Parameter λ

Output: Covariance Matrix W

$W = \hat{\Sigma} + \lambda I;$

while $\text{ave}(|W_{prev} - W|) > t \text{ ave}(|\text{offdiag}(\hat{\Sigma})|)$ **do**

for $i=1 \dots p$ **do**

 // Partition the matrix into all but the i th row and column
 and the i th row and column without the i th value

$W_{11} \leftarrow W[1 \dots p \neq i][1 \dots p \neq i];$
 $w_{12} \leftarrow W[1 \dots p \neq i];$
 $s_{12} \leftarrow \hat{\Sigma}[1 \dots p \neq i];$
 // Run coordinate descent to calculate $\hat{\beta}$

while $\text{norm}(\hat{\beta}_{prev} - \hat{\beta}) > t$ **do**

for $j=1 \dots p$ **do**

$\hat{\beta}_{prev} \leftarrow \hat{\beta};$
 $V \leftarrow W_{11};$
 $u \leftarrow s_{12};$
 $\hat{\beta}_j \leftarrow \frac{S_t(u_j - \sum_{k \neq j} V_{kj} \hat{\beta}_k, \lambda)}{V_{jj}};$

end

end

$W_{prev} \leftarrow W;$
 // Update W
 $W[1 \dots p \neq i] \leftarrow W_{11} \cdot \hat{\beta};$

end

end

$\text{offdiag}(M) = \text{offdiagonal elements of matrix } M$
 $S_t(x, t) = \text{sign}(x)(|x| - t)_+$
 $|M| = \text{absolute values of } M$
 $\text{ave}(M) = \text{average of } M$
 $\text{norm}(M) = L_2 \text{ norm of } M$

Fig. 1. Pseudo-code for the Graphical Lasso

2 Methods and Data

In this work, instead of the general Markowitz frontier, we focus on the minimum variance portfolio, unconstrained by the expected return (*i.e.* one corner of the frontier curve). This allows us to negate the errors from mean estimation in constructing our portfolios.

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \mathbf{w}^T \Sigma \mathbf{w} \\ & \text{subject to} && \sum_{i=1}^{i=N} w_i = 1 \\ & && w_i \geq 0 \end{aligned} \tag{8}$$

The covariance (Σ) is estimated using the empirical covariance and the glasso, with the convergence threshold (t) set to 0.001 and the regularisation parameter (λ) varied with the dataset, due to the differences in the means and variances of the datasets (monthly returns will in general have larger absolute values than daily returns). We use an implementation of the glasso written using Cython [3], and CVXPY [9] to solve the optimisation problem.

Two datasets are used to test our methods, with different sizes and time periods. The first is taken from Center for Research in Security Prices (CRSP) database from the 30th November 1982 until the 31st December 1990. We use the monthly percentage returns. Any assets with incomplete data are removed, leaving us with 92 observations from 26 companies. The company sectors are shown in Table 1. We set λ to 0.01 for this dataset. Our second dataset is the S&P500 daily percentage returns from 2nd January 2003 until 25th January 2007. Again, assets with incomplete data are removed, leaving us with 1259 observations from 409 companies. We set λ to 0.0002 for this dataset. Both are split into 2 equally sized sets, a training set to calculate the optimal portfolio from and a test set to evaluate the out of sample performance.

To test the robustness of the covariance estimation, we remove a number of samples randomly from the training data and calculate the mean and covariance from this corrupted data set. For the CRSP data we remove 4, and for the S&P500 data we remove 60, due to the much larger size of this dataset. We then solve the unconstrained risk minimization problem to get a set of portfolio weights. These weights are used to calculate a mean risk and return from the unseen data. This is run 25 times, although similar results are obtained when the number of runs is set to 10, 15 and 20. We can then compare how the corruption of the data affects the risks and returns of the portfolios produced on both seen and unseen data.

Following this we construct a network from the sparse precision matrix produced by the glasso. Focusing on the CRSP data due to the smaller number of companies, λ is set to 0.005. A link between companies implies their stocks are correlated in some way. Increasing the value of λ increases the sparsity of the precision matrix and will result in fewer links.

Table 1. Number of companies in each sector from the CRSP database. The colouring and number represents the sector they belong to in Figure 6

Sector	Number of Companies	Colour	Number
Agriculture	1	Brown	1
Communications	2	Cyan	2
Cyclical Consumer Goods & Services	5	Green	3
Energy	2	Red	4
Financials	2	Light Green	5
Industrials	1	Navy Blue	6
Insurance	2	Purple	7
Non-Cyclical Consumer Goods & Services	1	Grey	8
Technology	9	Yellow	9
Utilities	1	Orange	10

3 Results

In this section we test the portfolios constructed using the glasso covariance against those constructed using the empirical covariance and explore why they perform differently. Firstly we look at the CRSP data.

Figure 2 shows boxplots of the risk and return for the risk minimized portfolios using the CRSP data. We can see the portfolios produced using the glasso covariance have a comparable or slightly smaller variance and fewer outliers in their risk and returns on the training set, and a much smaller variance in their test set than those produced with the empirical covariance.

Figure 3 shows boxplots of the risks and returns for the S&P500 data. Again, we can see the significant reduction in variance of the risks and returns from using the glasso covariance. Due to room constraints, we do not show the training set boxplots, although these are similar to the results on the training set of CRSP data.

The results in Figure 2 are generated using a λ of 0.01. What happens if we change our choice of λ ? Figure 4 shows the portfolio weights chosen as we vary λ . As λ is increased, the portfolio weights become more evenly distributed throughout the companies. We would expect this to reduce the variance of risks and returns of the portfolios produced.

Now we explore relationships between the CRSP companies and how this affects our risk minimized portfolios. Figure 5 shows the generated network. Companies are tagged according to the sector the company belongs to. Sector colours and numbers are shown in Table 1. There is a general trend towards companies in the same sector being linked and unrelated companies being unlinked. Note in particular how 2 out of the 3 companies with no links are also the only members of their sector.

Relating this back to the optimal portfolios, we colour the nodes according to the weight put upon the company in Figure 6. The darker the node, the

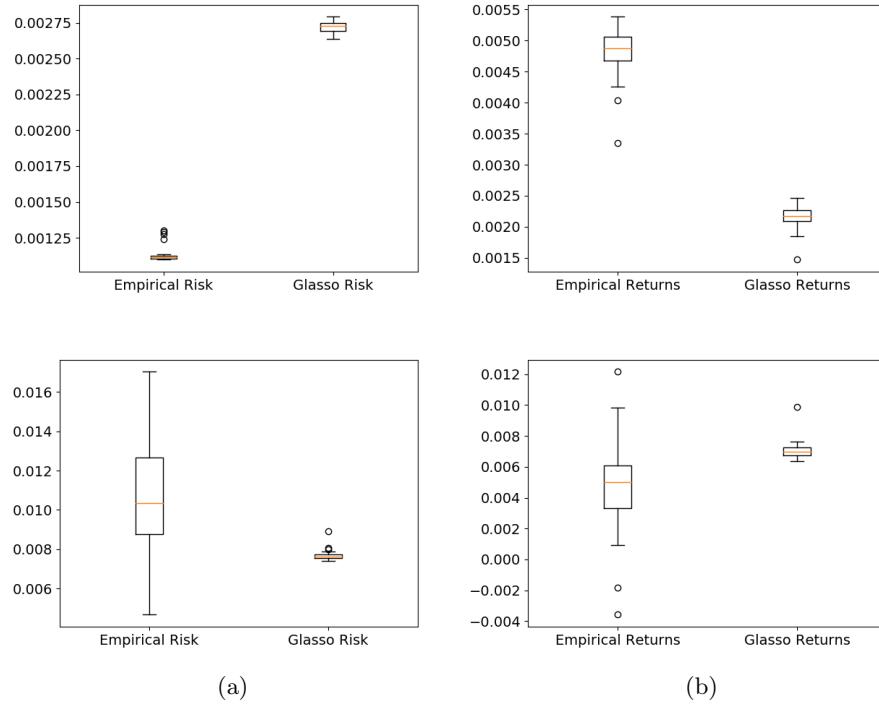


Fig. 2. Variation in risks (a) and returns (b) of portfolios on monthly returns of 26 companies using the empirical and glasso covariance. 4 samples are randomly removed each time (11.5% of the data) for 25 runs. The top row is the training set and the bottom row the test set. The portfolios produced using the glasso covariance have a slightly smaller variance and fewer outliers in their training risks and returns, and a much smaller variance in their test risks and returns.

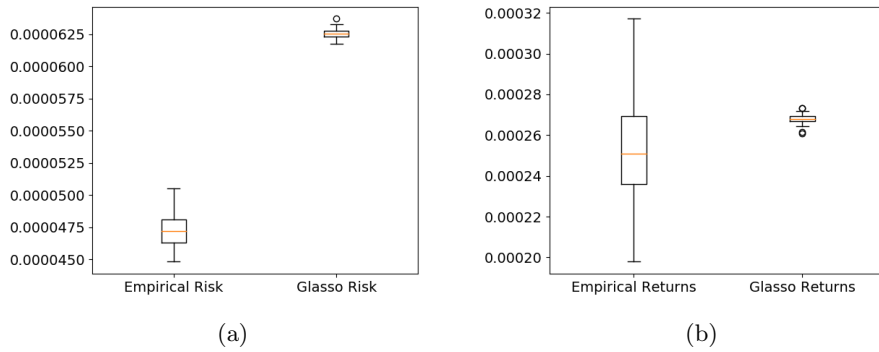


Fig. 3. Variation in risks (a) and returns (b) of portfolios on unseen daily returns of S&P500 companies. 60 samples (9.5% of the data) are removed each time. Again we can see the reduction in variance the glasso covariance provides

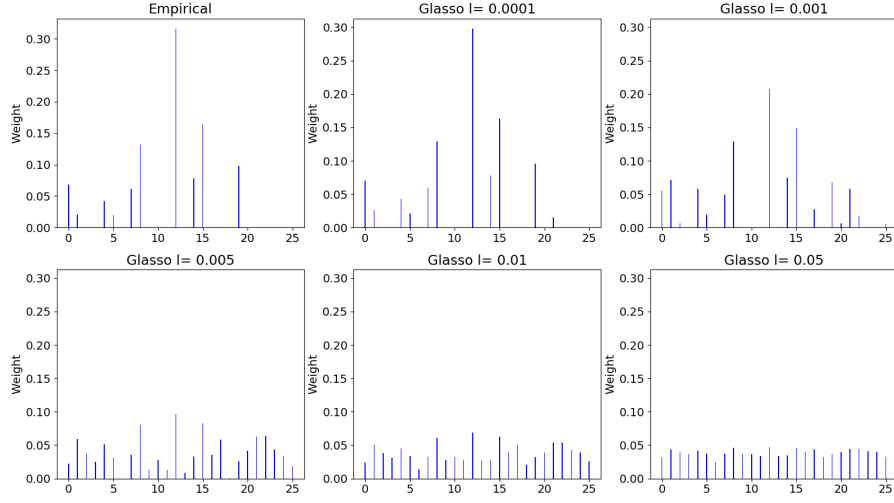


Fig. 4. Weights produced by the minimum risk unconstrained optimisation using empirical and glasso covariance. As we increase the regularization parameter for the graphical lasso, the weights tend to spread out to a $1/n$ portfolio.

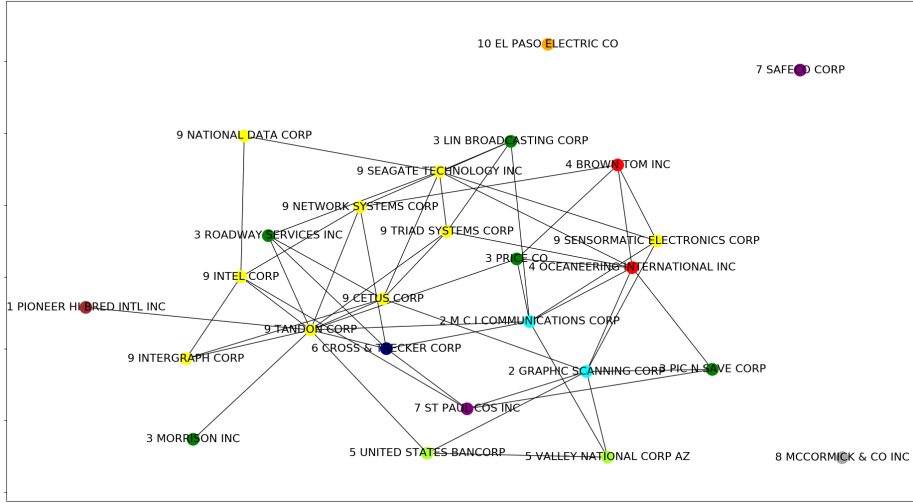


Fig. 5. Links between companies produced by the graphical lasso with λ set to 0.005. Companies are tagged according to the sector that they belong to - the number at the start of the label and the node colour indicate this (Colours and numbers are shown in Table 1). Note how companies in the same sectors are linked, particularly the cluster of technology companies.

larger the weight upon it. We can see that companies linked together have lower weights than those not linked. This should reduce the risk by diversifying the portfolio and avoiding companies who have correlated stocks.

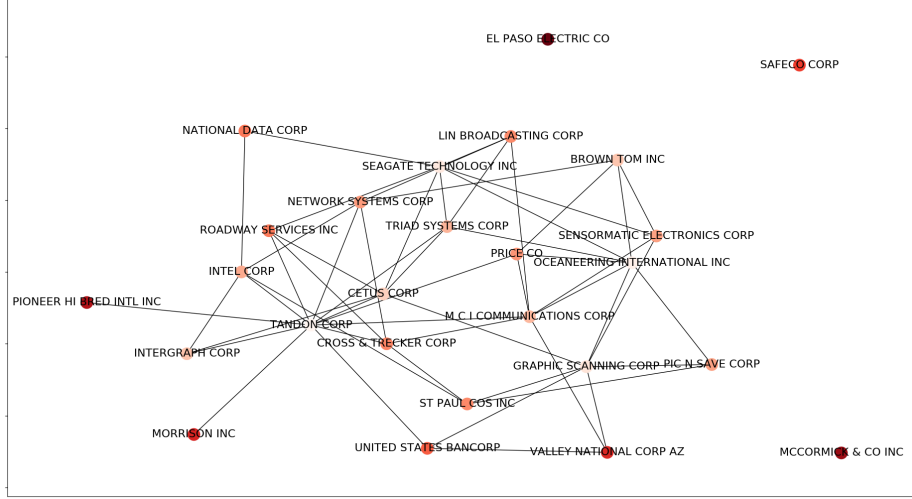


Fig. 6. Links between companies produced by the graphical lasso. The darker the node, the larger the weight on that company in the portfolio. The companies who are linked are lighter than those without links. This provides diversification away from correlated stocks and so should reduce the risk of the portfolio.

4 Conclusion

In this paper we have shown that using the graphical lasso to estimate covariance matrices can improve the robustness of portfolio optimization. The portfolios constructed using the graphical lasso have a lower variance of risks and returns than those constructed using the empirical covariance, particularly in their unseen data. We demonstrate this by removing data from the training set and comparing the risks and returns of the portfolios produced with both monthly and daily data. Finally we construct a network using the precision matrix estimated by the glasso to explore relationships between companies and how this affects the choice of assets to buy, with the minimum risk portfolios preferring to reduce their purchases of correlated stocks.

Acknowledgements This work was partially funded by the Engineering and Physical Sciences Research Council, UK (EP/N014189: Joining the Dots, from Data to Insight).

References

1. Awoye, O.A.: Markowitz Minimum Variance Portfolio Optimization using New Machine Learning Methods. Ph.D. thesis, (UCL) University College London (2016)
2. Banerjee, O., El Ghaoui, L., d’Aspremont, A.: Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *J. Mach. Learn. Res.* 9, 485–516 (2008)
3. Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D., Smith, K.: Cython: The best of both worlds. *Computing in Science Engineering* 13(2), 31–39 (2011)
4. Best, M.J., Grauer, R.R.: On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *Review of Financial Studies* 4(2), 315–342 (1991)
5. Boginski, V., Butenko, S., Pardalos, P.M.: Statistical analysis of financial networks. *Computational Statistics & Data Analysis* 48(2), 431–443 (2005)
6. Brodie, J., Daubechies, I., De Mol, C., Giannone, D., Loris, I.: Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences of the United States of America* 106(30), 12267–72 (2009)
7. Chu, M.T., Funderlic, R.E., Plemmons, R.J.: Structured low rank approximation. *Linear Algebra and its Applications* 366, 157–172 (2003), special issue on Structured Matrices: Analysis, Algorithms and Applications
8. DeMiguel, V., Garlappi, L., Uppal, R.: Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy? *The review of Financial studies* 22(5), 1915–1953 (2007)
9. Diamond, S., Boyd, S.: CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research* 17(83), 1–5 (2016)
10. Fan, J., Fan, Y., Lv, J.: High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147(1), 186–197 (2008), econometric modelling in finance and risk management: An overview
11. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9(3), 432–441 (2008)
12. Goto, S., Xu, Y.: Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis* 50(6), 1415–1441 (2015)
13. Huang, W.Q., Zhuang, X.T., Yao, S.: A network analysis of the chinese stock market. *Physica A: Statistical Mechanics and its Applications* 388(14), 2956–2964 (2009)
14. Ledoit, O., Wolf, M.: Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management* 30(4), 110–119 (2004)
15. Mantegna, R.N.: Hierarchical structure in financial markets. *The European Physical Journal B - Condensed Matter and Complex Systems* 11(1), 193–197 (1999)
16. Markovsky, I., Niranjan, M.: Approximate low-rank factorization with structured factors. *Computational Statistics & Data Analysis* 54(12), 3411–3420 (2010)
17. Markowitz, H.M.: Portfolio selection. *The Journal of Finance* 7(60), 77–91 (1952)
18. Qiu, H., Han, F., Liu, H., Caffo, B.: Robust portfolio optimization. In: *Proceedings of the 28th International Conference on Neural Information Processing Systems*. pp. 46–54. NIPS’15, MIT Press, Cambridge, MA, USA (2015)
19. Takeda, A., Niranjan, M., Gotoh, J.Y., Kawahara, Y.: Simultaneous pursuit of out-of-sample performance and sparsity in index tracking portfolios. *Computational Management Science* 10(1), 21–49 (2013)