Large Covariance Matrix Estimation With Auxiliary Information

Shuyi Ge, Shaoran Li, Oliver Linton, Weiguang Liu

March 1, 2021

Abstract

This paper considers incorporating auxiliary information about the location of significant correlations into the estimation of large-dimensional covariance matrices. The estimated covariance will emphasize economically meaningful links instead of achieving sparsity by pure statistical methods and it provides a way to combine new sets of information from machine learning and textual analysis with traditional datasets. An preliminary empirical study shows comparable performance with purely statistical methods in a portfolio management setting.

1 Model and Introduction

The goal is to estimate $\Sigma = \text{var}(y)$ where y is a $p \times 1$ random vector, say asset returns. We collect the T observations into a matrix $Y : p \times T$. Sample covariance estimate $\hat{\Sigma} = \frac{1}{T}(Y - \bar{Y}\mathbf{1})(Y - \bar{Y}\mathbf{1})'$ is problematic when p is not small relative to T. Popular estimation strategies include factor model, shrinkage, thresholding, banding, tapering, etc, see for example.

If in addition to the observaion of Y, we have observed an additional network G among the firms, where G_{ij} either takes value $0, 1$ or a score in $[0, 1]$, with higher G_{ij} implying that it's more "likely" that the returns of firm i, j are correlated. Then this auxiliary information can be used to help estimate the covaraince matrix Σ.

For example, such G information could come from textual analysis, that has become more and more popular in finance, for example (Fan, Xue, and Zhou, 2021). For example, Hoberg and Phillips, 2016 identifies a product similarity network from financial reports that has been shown to be more accurate than industry block diagonal matrix. As linked firms are potentially subject to similar demand shock, we have reason to believe that G contains valuable information about the comovement among the returns.

This paper aims to provide ways to extract the information contained in the auxiliary G matrix to help estimate the covaraince Σ. The two methods we have so far considered are

1. Guided Correlation Thresholding: we apply thresholding to the correlation matrix, where the threshold level depends either linearly on or is related to the network score in a probit model.

2. Guided Linear Shrinkage: we adopt the linear shrinkage method, where the shrainkge targets is chosen based on the network information. This is different from previous practice of shrinking to the identity or equicorrelation matrix.

More specifically, we consider the following *Adaptive Correlation Thresholding* method, suppose we observe Y_t for $t = 1, \ldots, T$, the procedure is

1. Estimate the sample covariance estimate $\hat{\Sigma}$, and the sample correlation matrix $\hat{R}$.

2. Apply the generalized thresholding function $h(r_{ij}, \tau_{ij})$ to the off-diagonal elements of $\hat{R} = (\hat{r}_{ij})$, Rothman, Levina, and Zhu, 2009. The novelty is now we allow the threshold τ_{ij} to vary across elements and to depend on the network information. Specifications we have considered for the threshold τ are

   - Simple linear model
   $$\tau(G_{ij}) = a + bG_{ij}$$

   - The probit model
   $$\tau_{ij} = \tau\big(G_{ij}\big) = \Phi\big(a + b\big|G_{ij}\big|\big)$$

3. Estimate the unknown parameters in the τ function by cross validation, as in Bickel and Levina, 2008, Cai and W. Liu, 2011, where we randomly split the sample V times, for each v, compute the new estimator $\hat{\Sigma}_G^{1,v}$ with the first subsample,

and sample covariance $\hat{\Sigma}^{2,v}$ and the criterion is

$$L(a,b) = \frac{1}{V}\sum_{v}^{V}\left\|\hat{\Sigma}_G^{1,v} - \hat{\Sigma}^{2,v}\right\|_F^2$$

we find a, b that minimise this criterion.

4. Then with this estimation of a, b, we can estimate Σ on the test sample.

For the second method we consider, we construct a linear shrinkage target based on the hard-threshold version of $\hat{\Sigma}$, call it $\hat{\Sigma}_H = [\hat{\Sigma}_{ij}\mathbf{1}\{G_{ij} > 0\}]$. And apply linear shriankge with the target.

There are several advantages of using network guided method:

1. The main advantage is that we can construct network that has economic meaning. Comparing to purely data-driven thresholding or shrinkage methods, the method provides more robustness and efficiency, if we can recover the "real" links from the network. The relationship identified from auxialiary network will be more stable over time than the relationship identified from return data alone.

2. In this paper ,we have used existing networks, we can include multiple network estimates in the τ, perhaps also characteristics of the company, so it's both flexible and extensible. It also provides a way to discern which set of information is relevant based an estimate of the coefficients a, b in the thresholding level.

3. The networks may provide industry-level comovement that is potentially related to the "weak factors" components, which we intend to investigate.

2 Literature Review

There has been extensive research on high-dimensional covariance estimation. Some important lines of thinking include elementwishe banding and thresholding method, shriankge method, factor models, etc. For a book-length review see Pourahmadi, 2013.

Bickel and Levina, 2008 considers banding or tapering the sample covariance matrix. Bickel and Levina, 2008 considers covariance regularization by hard thresholding. They also compare the results between banding when there is a natural ordering(for example, time series autocorrelation) and thresholding where we need to pay a $\log p$ price in the convergence rate to learn the locations. Cai and W. Liu, 2011's adptive thresholding where instead of doing T_{t_n}, threshold as:

$$\hat{\sigma}^*_{ij} = s_{t_{ij}}(\hat{\sigma}_{ij}) \tag{1}$$

where 1. $|s_\lambda(z)| \leq c|y|$ for all $|z-y| \leq \lambda$ 2. $s_\lambda(z) = 0$ for $|z| \leq \lambda$ 3. $|s_\lambda(z) - z| \leq \lambda$. The convergence rate is the same, although here the uniformity class is larger. Fan, Liao, and H. Liu, 2015 proposes thresholding on the correlation matrix. The choice of thresholding functions can be found Rothman, Levina, and Zhu, 2009, Fan and Li, 2001, etc.

As an application of thresholding method, Fan, Furger, and Xiu, 2016 use hard thresholding method in a high-frequency setting based on the sector/industry classifier. $s_{ij}(\sigma_{ij}) = \sigma_{ij}$ if ij are in the same industry. The network they is a block-diagonal matrix and our results accommodate more general and flexible network information.

Ledoit and Wolf, 2004 develops an estimation strategy based on linear shrinkage, where the target is identity matrix. This shrinkage guarantees that the estimated covariance matrix is well-conditioned. This approach can be thought of as decreasing varaince at the expense of increasing bias a little. There are articles discuss multiple targets, for example, Schäfer and Strimmer, 2005, Lancewicki and Aladjem, 2014 and Gray et al., 2018, but their targets are either fixed or data-driven, so different from our guided method where we bring in new information from auxiliary network information. Ledoit and Wolf, 2012 and Ledoit and Wolf, 2017 propose nonlinear shrinakage where the eigenvalues are pulled towards the "correct level" solving a nonrandom limit loss function. The shrinkage method has been shown to have really good performance in estimating large-dimensional covaraince matrix, however they are a global method whereas our method is designed to emphasize "economically meaning" links. There is also a vast literature on factor models in high-dimensional models and applications in empirical finance. We refer to Connor, Hagmann, and Linton, 2012, Fan, Liao, and H. Liu, 2015 and Fan, Liao, and Wang, 2016 and literature review therein.

3 Simulation

We generate $T = 200$ independent samples from $p = 500$ dimensional normal distribution $N(0, \Sigma_1)$, where Σ_1 is constructed by concatenating four AR(1) serial correlation matrix with ones on the diagonal and parameter $\rho = 0.9$. The network $G = [G_{ij}]$ where $G_{ij} = 1$ if $\sigma_{ij} > 0.6$.

Given the sample X and the G matrix, we compute the correlation matrix R and apply adaptive correlation thresholding $h(r_{ij}, \tau_{ij})$ to the off-diagonal elements of R. The τ_{ij} depends on G_{ij} and is specified here as probit function so that it's between $0, 1$

$$\tau_{ij} = \Phi(a + bG_{ij})$$

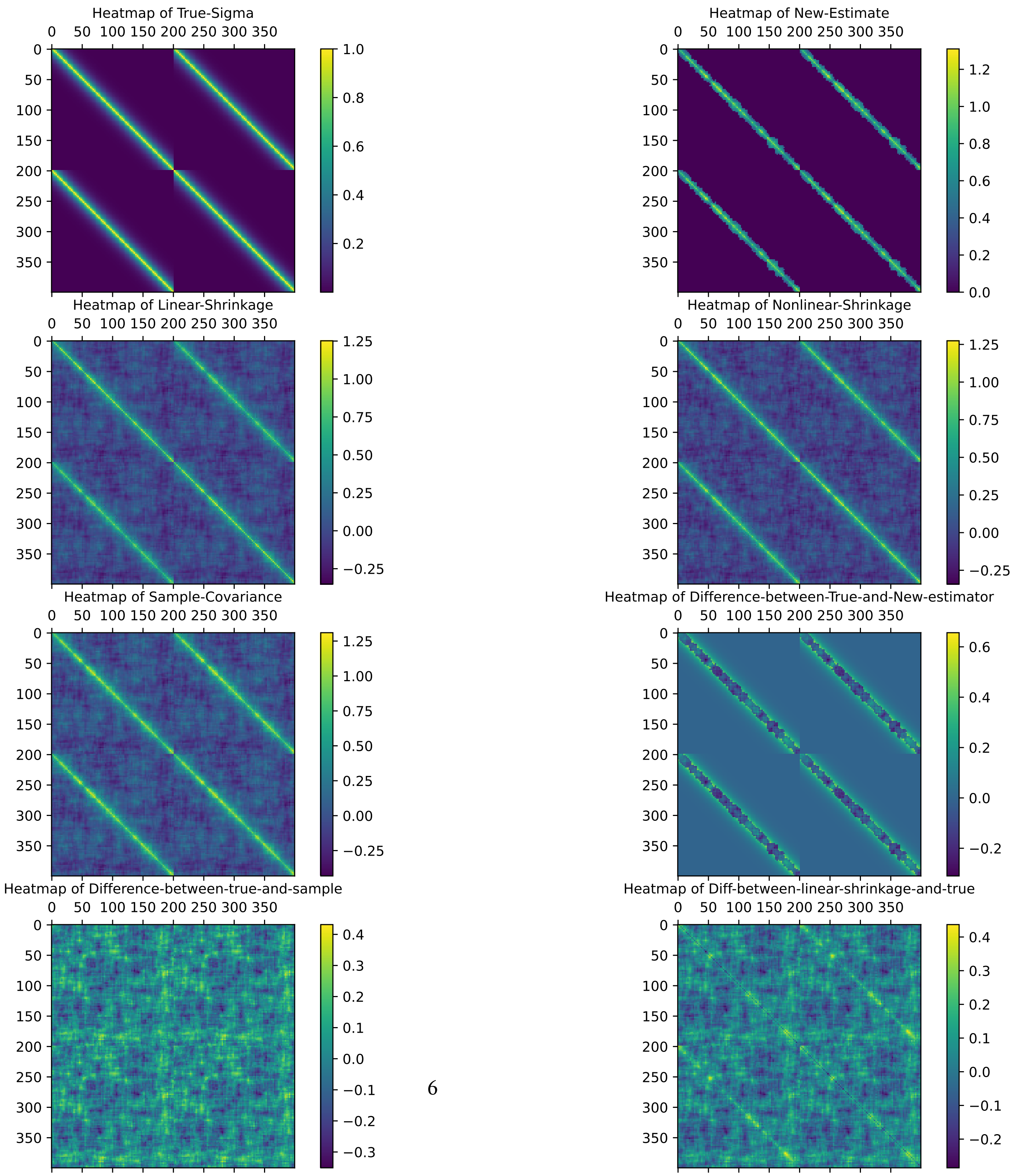
The following Table 1 describes the Frobenius-norm error using different estimators, the new estimator which is optimized to minimise the Frobenius error, has good performance. In Figure 1 we show the heatmaps of the true Σ and the estimates. Perhaps using a softer thresholding function will give better result. 2. We are trying more meaningful specification of Σ and we suspect that the new estimator will be good at preserving important structural features of the covariance.

New estimator	35.5759973173453
Linear shrinkage	36.44438551876532
Nonlinear shrinkage	35.09953816084891
Sample	40.62939042523703
Sample - New	49.10202294326661

Table 1: Difference with true Σ in terms of Frobenius Norm

4 Empirical Study

In this section, we apply the adaptive correlation thresholding method to a portfolio construction problem. First we describe the procedure and then present some of the results we have.


Heatmap of True-Sigma
Heatmap of New-Estimate
Heatmap of Linear-Shrinkage
Heatmap of Nonlinear-Shrinkage
Heatmap of Sample-Covariance
Heatmap of Difference-between-True-and-New-estimator
Heatmap of Difference-between-true-and-sample
Heatmap of Diff-between-linear-shrinkage-and-true
0 50 100 150 200 250 300 350

4.1 Covariance Estimation Procedure

Assume we observe the excess return Y_{it}, $i = 1, \ldots, N$ and $t = 1, \ldots, T$ follows

$$Y_{it} = B_i' F_t + u_{it}; \quad \Sigma_u = E(u_t u_t')$$

where F_t are factor excess returns. Here we have considered Fama-French 3 and the Carhart's momentum factor. The goal is to estimate $\Sigma_Y = E(YY')$ and use that estimate to construct portfolio following Ledoit and Wolf, 2017. The auxialiary network we have include the Hoberg and Phillips, 2016's network(henceforth Hoberg's Network) and IBES analysts cocoverage network. Here we present the results for SP500 returns using Hoberg's Network.

The procedure we take is as follows.

1. We run time series linear regressions of Y_{it} on $F_{k,t}$, obtain the beta estimates $\hat{B}_i$ and the residual $\hat{u}_{it}$.

2. Compute the covaraince matrix $S_{\hat{u}} = \frac{1}{T}\hat{u}\hat{u}'$ and $S_F = \frac{1}{T}\sum_t (F_t - \bar{F})(F_t - \bar{F})'$ and appply *adaptive correlation thresholding* on $S_{\hat{u}}$, denote the estimate as $\hat{S}_{\hat{u}}$. where the second step adaptive correlation thresholding is achieved in the following way. Let R_u be the correlation matrix calculated from S_u. We use soft thresholding $h(r_{ij}, \tau_{ij}) = \text{sign}\left(r_{ij}\right)\left(r_{ij} - \tau_{ij}\right)_+$ on the off-diagonal elemetns r_{ij} of R_u, where

$$\tau_{ij} = \delta_{ij}\sqrt{\frac{\log N}{T}}$$

and

$$\delta_{ij} = a + bG_{ij}$$

Let the threshold estimate be $\hat{R}_{\hat{u}}(a, b)$, given a, b, our estimate will be

$$\hat{S}_{\hat{u}} = \hat{S}_{\hat{u}}(a, b) = \text{diag}(S_{\hat{u}})^{\frac{1}{2}} \hat{R}_{\hat{u}} \, \text{diag}(S_{\hat{u}})^{\frac{1}{2}}$$

In order to guarantee positive definiteness, I follow the suggestion in Fan, Liao, and H. Liu, 2015 and Fan, Liao, and Mincheva, 2013, by first finding the minimum $\underline{\delta}$ such that the $\hat{S}(\delta, 0)$ has its smallest eigenvalue larger than 0 if we choose

$\tau_{ij} = \underline{\delta}\sqrt{\frac{\log N}{T}}$.

Then a, b are estimated using cross-validation following Bickel and Levina, 2008 by randomly spliting the sample V times, for each $v = 1, \ldots, V$, compute the estimate $\hat{S}_u^{1,v}$ with the first subsample, and sample covariance estimate $\hat{\Sigma}_u^{2,v}$ with the second subsample and let the criterion function be

$$L(a, b) = \frac{1}{V}\sum_{v}^{V}\left\|\hat{S}_u^{1,v} - \hat{\Sigma}_u^{2,v}\right\|_F^2$$

we find $\hat{a}, \hat{b}$ that minimise this criterion subject to the constraints:

$$0 \leq a\sqrt{\frac{\log N}{T}} \leq 1 \tag{2}$$

$$b\sqrt{\frac{\log N}{T}} \leq 0 \tag{3}$$

$$\underline{\delta} \leq a + b \tag{4}$$

3. Construct an estimate of Σ_Y by $\hat{\Sigma}_Y = \hat{B}S_F\hat{B}' + \hat{S}_{\hat{u}}$

4.2 Empirical Study

We have done a portfolio construction similar to that considered in Ledoit and Wolf, 2004 and Ledoit and Wolf, 2017. We collect daily data on SP500 stock returns R_{it} from 2006 to the end of 2017; FF3 factor returns $F_{kt}, k = 1, 2, 3$ and risk free rates $R_{f,t}$; Hoberg's network score matrices G_t that are updated yearly. $Y_{it} = R_{it}R_{f,t}$.

The Hoberg's Network is a yearly updated $N \times N$ network with elements in $[0, 1]$ with higher score G_{ij} reflecting potentially higher correlation between the i-th and j-th firms.

We divide the sample into windows of an estimation periods of 252 days followed by 21-day testing period and then these two windows are shifted by 21 days. In the estimation periods in window $m = 1, \ldots, M$ we construct estimate $\hat{\Sigma}_{Y,m} = \hat{B}S_F\hat{B}' + \hat{S}_{\hat{u}}\left(\hat{a}_m, \hat{b}_m\right)$ and in the testing periods, we use this estimate to construct a portfolio with weights in a vector $w_m = \frac{\mathbf{1}'\hat{\Sigma}_{Y,m}}{\mathbf{1}'\hat{\Sigma}_{Y,m}\mathbf{1}}$ and we record the profolio return over the testing period $w_m R_t$ for t in the testing period of window m. In this way, we obtain a daily time series of portfolio returns and the portfolio is updated roughly monthly.

Then we will have a time series of daily portfolio returns from 2007 to 2017 and we compute the standard deviation, average return and the Sharpe ratio. The results are in Table 2

	Standard deviation	Average Return × 250	Sharpe Ratio
Adaptive Correlation Thresholding	0.00728	0.0988*	0.0542*
Linear Shrinkage	0.00757	0.0868	0.0458
Nonlinear Shrinkage	0.00695*	0.0906	0.0521

Table 2: Comparison between different estimates

The new estimator is superior to linear shrinkage estimator in all three aspects, and only loses to nonlinear shrinkage in terms of standard deviation.

References

Bickel, Peter J. and Elizaveta Levina (2008). "Covariance Regularization by Thresholding". In: *The Annals of Statistics* 6, pp. 2577–2604 (cit. on pp. 2, 3, 8).

Cai, Tony and Weidong Liu (2011). "Adaptive Thresholding for Sparse Covariance Matrix Estimation". In: *Journal of the American Statistical Association* 494, pp. 672–684 (cit. on pp. 2, 4).

Connor, Gregory, Matthias Hagmann, and Oliver Linton (2012). "Efficient Semiparametric Estimation of the Fama–French Model and Extensions". In: *Econometrica* 2, pp. 713–754. ISSN: 1468-0262. DOI: 10.3982/ECTA7432 (cit. on p. 4).

Fan, Jianqing, Alex Furger, and Dacheng Xiu (Sept. 15, 2016). "Incorporating Global Industrial Classification Standard Into Portfolio Allocation: A Simple Factor-Based Large Covariance Matrix Estimator With High-Frequency Data". In: *Journal of Business & Economic Statistics*. ISSN: 0735-0015. URL: https://www.tandfonline.com/doi/pdf/10.1080/07350015.2015.1052458?casatoken=FtSi9R7i-SkAAAAA:VJxD-tO9DDtdi7QcU4Adch4gVpy6SXPZ0N9zWAyL0yXC8aAm7H9HT7NieTYfbdn21OpcJP5boOA (cit. on p. 4).

Fan, Jianqing and Runze Li (2001). “Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties”. In: *Journal of the American statistical Association* 456, pp. 1348–1360 (cit. on p. 4).

Fan, Jianqing, Yuan Liao, and Han Liu (Apr. 12, 2015). *An Overview on the Estimation of Large Covariance and Precision Matrices*. arXiv: 1504.02995 [stat]. URL: http://arxiv.org/abs/1504.02995 (cit. on pp. 4, 7).

Fan, Jianqing, Yuan Liao, and Martina Mincheva (2013). “Large Covariance Estimation by Thresholding Principal Orthogonal Complements”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 4, pp. 603–680. ISSN: 1467-9868. DOI: 10.1111/rssb.12016 (cit. on p. 7).

Fan, Jianqing, Yuan Liao, and Weichen Wang (Feb. 2016). “Projected Principal Component Analysis in Factor Models”. In: *Annals of Statistics* 1, pp. 219–254. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/15-AOS1364 (cit. on p. 4).

Fan, Jianqing, Lirong Xue, and Yang Zhou (Jan. 11, 2021). *How Much Can Machines Learn Finance From Chinese Text Data?* SSRN Scholarly Paper ID 3765862. Rochester, NY: Social Science Research Network. URL: https://papers.ssrn.com/abstract=3765862 (cit. on p. 1).

Gray, Harry et al. (Sept. 21, 2018). *Shrinkage Estimation of Large Covariance Matrices Using Multiple Shrinkage Targets*. arXiv: 1809.08024 [stat]. URL: http://arxiv.org/abs/1809.08024 (cit. on p. 4).

Hoberg, Gerard and Gordon Phillips (2016). “Text-Based Network Industries and Endogenous Product Differentiation”. In: *journal of political economy*, p. 43 (cit. on pp. 2, 7).

Lancewicki, Tomer and Mayer Aladjem (Dec. 2014). “Multi-Target Shrinkage Estimation for Covariance Matrices”. In: *IEEE Transactions on Signal Processing* 24, pp. 6380–6390. ISSN: 1941-0476. DOI: 10.1109/TSP.2014.2364784 (cit. on p. 4).

Ledoit, Olivier and Michael Wolf (July 31, 2004). “Honey, I Shrunk the Sample Covariance Matrix”. In: *The Journal of Portfolio Management* 4, pp. 110–119. ISSN: 0095-4918, 2168-8656. DOI: 10.3905/jpm.2004.110 (cit. on pp. 4, 8).

— (Apr. 2012). “Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices”. In: *Annals of Statistics* 2, pp. 1024–1060. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/12-AOS989 (cit. on p. 4).

Ledoit, Olivier and Michael Wolf (Dec. 1, 2017). "Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks". In: *The Review of Financial Studies* 12, pp. 4349–4388. ISSN: 0893-9454, 1465-7368. DOI: 10.1093/rfs/hhx052 (cit. on pp. 4, 7, 8).

Pourahmadi, Mohsen (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data*. John Wiley & Sons (cit. on p. 3).

Rothman, Adam J., Elizaveta Levina, and Ji Zhu (2009). "Generalized Thresholding of Large Covariance Matrices". In: *Journal of the American Statistical Association* 485, pp. 177–186 (cit. on pp. 2, 4).

Schäfer, Juliane and Korbinian Strimmer (2005). "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics". In: p. 32 (cit. on p. 4).