Augment Large Covariance Matrix Estimation With Auxiliary Network Information

Shuyi Ge[*], Shaoran Li[†], Oliver Linton[‡] and Weiguang Liu[§]

Faculty of Economics, University of Cambridge

April 30, 2021

Abstract

This paper aims to incorporate auxiliary information about the location of significant correlations into the estimation of high-dimensional covariance matrices. With the development of machine learning techniques such as textual analysis, granular linkage information among firms that used to be notoriously hard to get are now becoming available to researchers. Our proposed method provides an avenue for combining those auxiliary network information with traditional economic datasets to improve the estimation of a large covariance matrix. Simulation results show that the proposed adaptive correlation thresholding method generally performs better in the estimation of covariance matrices than previous methods, especially when the true covariance matrix is sparse and the auxiliary network contains genuine information. As a preliminary application, we apply the method to the estimation of the covariance matrix of asset returns. There are several extensions and improvements that we are considering.

[*]Author email:sg751@cam.ac.uk
[†]Author email:sl736@cam.ac.uk
[‡]Author email:obl20@cam.ac.uk
[§]Author email:wl342@cam.ac.uk

1 Model and Introduction

Our goal is to estimate $\Sigma = \mathrm{var}(y)$ 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.

If in addition to the observation of Y, we observe a 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. We show that this auxiliary network can be used to improve the estimation the covariance matrix Σ. Examples of such network include Hoberg and Phillips, 2016, who 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. Israelsen, 2016 and Kaustia and Rantala, 2020 both find that companies covered by the same analysts show similarities in many unobserved dimensions, and this analyst-based network could explain excess co-movement on top of common factors. With the development of machine learning techniques such as textual analysis, we are better at acquiring information from big data. Granular linkage information among firms that used to be notoriously hard to get due to its proprietary properties, now are becoming available to researchers. The question is, how to use those auxiliary network information to better estimate the co-movement between assets?

This paper aims to provide ways to extract the information contained in the auxiliary G matrix to help estimate the covariance Σ. We consider an *Adaptive Correlation Thresholding* method, where we apply thresholding to the correlation matrix, with the threshold level depending on network information. More specifically, 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 $\hat{R} = (\hat{r}_{ij})$, as in 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

2

information. Specifications we have considered for the threshold τ are

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

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

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 the estimates of a, b, we can estimate Σ on the test sample.

There are several advantages of using network guided method:

1. The main advantage is that we are combining economically meaningful network with market-based performance data. Comparing to purely data-driven thresholding or shrinkage methods, the method utilizes valuable information embedded in external network data, which provides more robustness and efficiency. aif our auxiliary network contains the "real" links from the network. The relationship identified will be more stable over time than the relationship identified from return data alone.

2. This method is very flexible and extensible. Although in our current analysis we only use one of the existing networks as our proxy for G, you are free to include many candidate networks in the τ. You may want want to include characteristics-based distances, as it has been documented that companies with similar characteristics exhibit additional co-movement on top of common risk factors (see Fernandez, 2011 for example). 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 element-wise banding and thresholding method, shrinkage 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 considers adptive thresholding where threshold takes the form:

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

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 use 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

variance 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 shrinkage 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 covariance 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 X from $N = 500$ dimensional normal distribution $N(0, \Sigma)$, where $\Sigma_{ij} = \rho^{|i-j|}$. We consider the following two types of generated networks,

1. $G_1 = [G_{ij}]$ where $G_{ij} = 1$ if $\Sigma_{ij} > l$, so that we observe $G_{ij} = 1$ if Σ_{ij} is larger than a threshold level l

2. $G_{2,ij} = \Sigma_{ij} + z_{ij}$, for $j > i$, where z_{ij} are independently drawn from $N(0, \sigma^2)$, so that $G_{2,ij}$ provides a noisy observation of Σ. We simulate with varying parameters l and σ.

The following four tables, shows the simulation results. The new adaptive correlation thresholding estimator performs better than shrinkage methods and simple sample covariance in terms of both the matrix 1-norm and Frobenius norm of the estimation errors, when the G provides more accurate information about the true covariance matrices (when l and σ is smaller); and also when the true covariance Σ is more sparse, the adaptive correlation thresholding seems to perform better.

Table 1: The Frobenius norm of the true Σ and the Frobenius norm of the difference between the estimates and Σ. We observe G_1 which is $\mathbf{1}\{\Sigma_{ij} \geq l\}$

Observe G_1		Σ	Sample Covariance	Linear Shrinakage	Nonlinear Shrinakage	Adaptive Corr Thresholding
ρ	l					
0.80	0.5	47.62	34.32	27.12	26.63	15.13
	0.6	47.62	35.16	27.29	26.71	14.18
	0.7	47.62	35.95	26.93	26.34	14.14
	0.8	47.62	35.97	27.20	26.58	14.82
	0.9	47.62	35.10	27.33	26.79	16.65
0.90	0.5	68.69	35.85	30.54	29.34	20.49
	0.6	68.69	35.08	32.01	30.99	30.08
	0.7	68.69	35.58	31.00	29.84	25.66
	0.8	68.69	36.04	31.31	30.23	20.94
	0.9	68.69	36.50	31.71	30.50	24.82
0.95	0.5	97.81	35.31	32.77	31.48	34.57
	0.6	97.81	36.22	33.56	32.58	33.83
	0.7	97.81	36.75	34.22	33.01	32.12
	0.8	97.81	37.22	34.23	33.03	40.94
	0.9	97.81	38.67	35.38	34.12	29.11
0.99	0.5	211.66	40.78	44.90	44.26	85.67
	0.6	211.66	41.79	39.84	41.41	57.07
	0.7	211.66	34.65	36.47	40.24	62.38
	0.8	211.66	41.71	41.20	42.50	77.88
	0.9	211.66	39.22	33.44	38.06	47.69

Table 2: The Frobenius norm of the true Σ and the Frobenius norm of the difference between the estimates and Σ.

Observe G_2		Σ	Sample Covariance	Linear Shrinakage	Nonlinear Shrinakage	Adaptive Corr Thresholding
ρ	σ					
0.80	0.0	47.62	35.64	27.36	26.73	15.46
	0.1	47.62	35.99	26.99	26.41	15.61
	0.2	47.62	35.87	27.06	26.42	14.60
	0.3	47.62	35.55	27.20	26.59	14.84
	0.4	47.62	35.89	27.46	26.82	17.21
0.90	0.0	68.69	36.93	31.73	30.38	28.81
	0.1	68.69	35.86	32.29	30.97	22.32
	0.2	68.69	36.09	31.73	30.65	35.13
	0.3	68.69	34.65	30.97	30.06	20.21
	0.4	68.69	34.61	30.94	30.15	23.54
0.95	0.0	97.81	37.62	33.57	32.14	25.91
	0.1	97.81	37.20	33.72	33.17	27.94
	0.2	97.81	33.57	30.47	29.91	25.93
	0.3	97.81	34.82	31.75	31.04	25.30
	0.4	97.81	34.79	33.65	32.82	33.42
0.99	0.0	211.66	31.72	30.93	32.38	68.24
	0.1	211.66	36.74	40.73	38.74	73.51
	0.2	211.66	40.72	36.44	38.85	51.01
	0.3	211.66	29.65	30.13	29.98	72.43
	0.4	211.66	47.19	41.00	45.82	120.97

Table 3: The matrix 1 norm of the true Σ and the matrix 1 norm of the difference between the estimates and Σ. We observe G_1 which is $\mathbf{1}\{\Sigma_{ij} \geq l\}$

Observe G_1		Σ	Sample Covariance	Linear Shrinakage	Nonlinear Shrinakage	Adaptive Corr Thresholding
ρ	l					
0.80	0.0	47.62	35.68	22.47	22.44	17.73
	0.1	47.62	35.12	22.29	22.54	14.82
	0.2	47.62	37.76	22.90	22.03	17.45
	0.3	47.62	34.04	21.65	22.14	15.26
	0.4	47.62	35.84	22.97	22.16	14.76
0.90	0.0	68.69	45.61	33.59	30.95	17.91
	0.1	68.69	39.91	30.26	29.55	20.97
	0.2	68.69	41.63	30.64	30.02	22.20
	0.3	68.69	37.60	30.26	28.93	20.29
	0.4	68.69	37.34	29.15	28.04	23.21
0.95	0.0	97.81	43.60	42.74	43.00	47.35
	0.1	97.81	39.52	37.92	36.84	31.95
	0.2	97.81	42.50	41.81	38.80	49.32
	0.3	97.81	45.70	40.13	40.16	30.88
	0.4	97.81	43.21	37.86	34.72	28.48
0.99	0.0	211.66	81.33	72.85	77.99	35.73
	0.1	211.66	50.42	43.82	53.34	48.87
	0.2	211.66	52.14	55.42	54.08	80.27
	0.3	211.66	66.11	58.51	60.91	51.53
	0.4	211.66	53.18	45.68	48.64	48.46

Table 4: The matrix 1 norm of the true Σ and the matrix 1 norm of the difference between the estimates and Σ.

Observe G_2		Σ	Sample Covariance	Linear Shrinakage	Nonlinear Shrinakage	Adaptive Corr Thresholding
ρ	σ					
0.80	0.5	47.62	38.02	23.62	22.64	15.23
	0.6	47.62	35.85	22.57	22.34	14.76
	0.7	47.62	35.75	22.06	22.18	15.11
	0.8	47.62	38.63	24.90	23.36	14.23
	0.9	47.62	36.34	21.73	21.83	15.35
0.90	0.5	68.69	37.90	30.23	28.42	22.30
	0.6	68.69	44.40	30.43	29.72	38.01
	0.7	68.69	37.73	30.74	29.48	21.02
	0.8	68.69	38.40	30.02	29.34	19.74
	0.9	68.69	38.24	31.76	29.07	19.73
0.95	0.5	97.81	37.95	36.71	36.22	29.38
	0.6	97.81	45.55	38.66	36.65	31.78
	0.7	97.81	39.66	35.85	35.35	29.67
	0.8	97.81	39.70	39.31	38.31	32.88
	0.9	97.81	50.46	39.55	38.01	41.03
0.99	0.5	211.66	50.51	47.20	49.90	67.03
	0.6	211.66	43.32	47.98	48.10	67.31
	0.7	211.66	43.90	40.44	41.07	67.02
	0.8	211.66	64.95	69.00	70.53	70.13
	0.9	211.66	42.75	45.67	45.99	71.32

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.

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 auxiliary 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 $h(r_{ij}, \tau_{ij}) = \operatorname{sign}\left(r_{ij}\right)\left(r_{ij} - \tau_{ij}\right)_+$ on the off-diagonal elemetns r_{ij} of R_u, where

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

and

$$\delta_{ij} = a + b G_{ij}$$

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

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

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 splitting 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:

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

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

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

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

We have estimated the covariance matrices of SP500 stocks from 1996 to the end of 2017; using stock return data and Fama-French 3 factor returns $F_{kt}, k = 1, 2, 3$. We incorporate Hoberg's network G_t that are updated yearly into our estimation procedure.

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.

In Figure 1, we present the distribution, we present the distribution (blue) of sample covariance estimates $S_{\hat{u},ij}$ of residuals after regressing the the stocks returns on the Fama-French 3 factor for the stocks that haIn Figure 1, we present the distribution (blue) of sample covariance estimates $S_{\hat{u},ij}$ of residuals after regressing the the stocks returns on the Fama-French 3 factor for the stocks tsiduals after regressing the the

stocks returns on the Fama-French 3 factor for the stocks that have no missing data in the dataset; alongwith the distribution of sample covariance $S_{\hat{u},ij}$ for those ij with $G_{ij} > 0$ in the Hoberg's network. It's clear that the distribution is shifted to the right, implying that Hoberg's network can pick up information that are not explained by the factors.
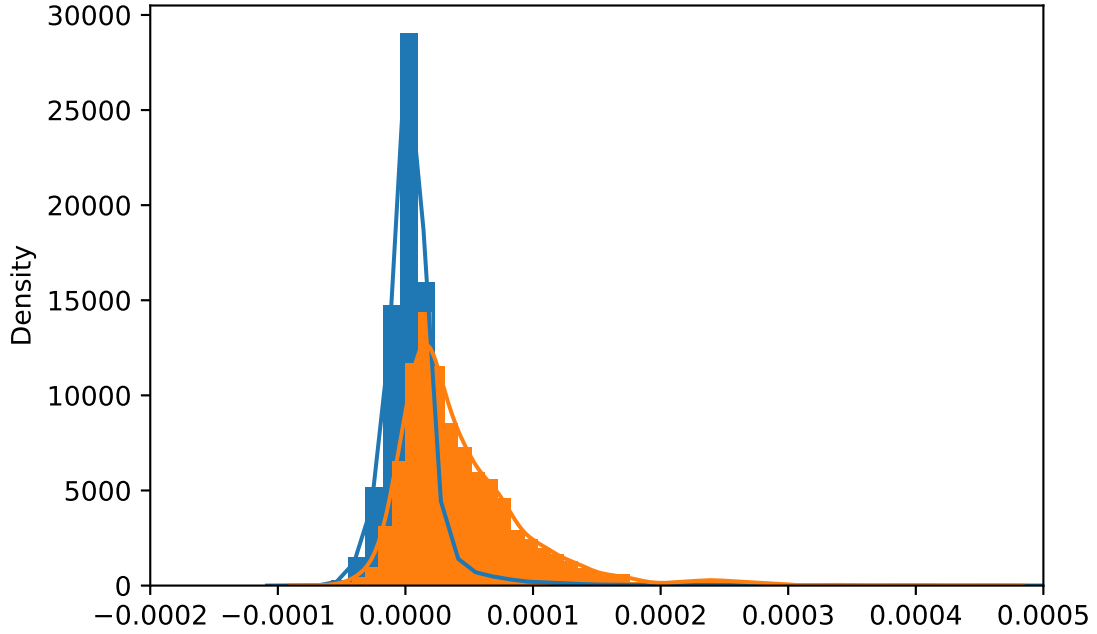

Figure 1: Distribution of $S_{\hat{u},ij}$, $i, j = 1, \ldots, N$ and $S_{\hat{u},ij}$ for i, j such that $G_{ij} > 0$

Then we use a rolling-window estimation, with 252-day estimation period and then move the window forward by 21 days. In the estimation periods in window $m = 1, \ldots, M$ we construct estimate $\hat{\Sigma}_{Y,m} = \hat{B}S_F\hat{B}' + \hat{S}_{\hat{u}}\left(\hat{a}_m, \hat{b}_m\right)$. The estimated parameters $\hat{a}_m, \hat{b}_m$ have mean $(1.177, -0.252)$, where the $\hat{b}$ measures the effect of knowing the auxiliary network on the thresholding level.

5   Conclusion and Further Works

This paper considers the problem of incorporating ever-increasing auxialiary data from machine learning techniques such as textual analysis into the estimation of large covariance matrices. This current version is preliminary with ongoing research on the following applications.

Firstly, we are applying the covaraince estimation technique on portfolio construction, following the problem considered in Ledoit and Wolf, 2004 and Ledoit and Wolf, 2017, where the estimation of the sparse covaraince matrices are vital for constructing the minimum-variance portfolio.

Secondly, the method can be applied to study spatial-APT under large N case. **kou2018asset** finds that common risk factors are insufficient to capture all the significant inter-dependencies in asset returns, and local interactions are also important. Spatial-APT and spatial CAPM type of models have not been popular in large N case since the measure of contiguity is challenging. Our method can uncover contemporaneously correlated entities by combining market-based information and auxiliary network information, thus providing a natural contiguity measure. Relying solely on either statistical methods or external network information is not as desirable as the links identified by the former are hard to interpret and the external network may miss some important links.

Thirdly, we are expanding the set of auxialiary networks beyond the Hoberg's network as well as applying the technique on larger datasets. We have collected IBES analysts cocovarage network and are constructing new network based on firms' characteristics. The flexibility of the methods allows us many potential improvements.

References

Bickel, Peter J. and Elizaveta Levina (2008). "Covariance Regularization by Thresholding". In: *The Annals of Statistics* 6, pp. 2577–2604 (cit. on pp. 3, 4, 11).

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. 3, 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. 5).

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-tO9DDtdi7QcU4Adch4gVpy6SXPZ0N9zWAyL0yXC8aAm7H9HT7NieTYfbdn210pcJP5boOA (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, 5, 11).

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. 11).

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. 5).

Fernandez, Viviana (2011). "Spatial linkages in international financial markets". In: *Quantitative Finance* 11.2, pp. 237–245 (cit. on p. 3).

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. 5).

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, 10).

Israelsen, Ryan D (2016). "Does common analyst coverage explain excess comovement?" In: *Journal of Financial and Quantitative Analysis*, pp. 1193–1229 (cit. on p. 2).

Kaustia, Markku and Ville Rantala (2020). "Common Analysts: Method for Defining Peer Firms". In: *Journal of Financial and Quantitative Analysis*, pp. 1–32. ISSN: 0022-1090, 1756-6916. DOI: 10.1017/S0022109020000514 (cit. on p. 2).

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. 5).

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, 13).

— (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. 5).

— (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. 5, 10, 13).

Pourahmadi, Mohsen (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data.* John Wiley & Sons (cit. on p. 4).

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. 5).