Augment Large Covariance Matrix Estimation With Auxiliary Network Information

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

Faculty of Economics, University of Cambridge

June 14, 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 $\Sigma = \text{var}(y)$ 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.

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

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 Introduction

Covariance matrix estimation is an important area of research in both finance and economics. Suppose we have observations $\mathbf{X}_t = (X_{1t}, \ldots, X_{Nt})^T$, $t = 1, \ldots, T$, that are from an $N-$dimensional random vectors with $Cov(\mathbf{X}) = E(\mathbf{X}\mathbf{X}^T) = \Sigma_X$. The most straightforward estimator is the sample covariance matrix, which is defined as follows:

$$\hat{\Sigma}_X = \frac{1}{T-1}(X - \bar{X})(X - \bar{X})^\intercal = [\sigma_{ij}]_{N\times N}, \tag{1}$$

where X is the $N \times T$ matrix of observations; $\bar{X} = \frac{1}{T}X1_T1_T^\intercal$, with 1_T being a $T \times 1$ vectors of 1. However, in the high-dimensional case, where the dimension N goes to infinity at a faster rate than that of the sample size T, the naive sample covariance matrix is ill-conditioned. Some structures need to be imposed on Σ_X, and regularisation techniques need to be applied to make sure the estimator is reliable. In particular, we usually assume that Σ_X is sparse (i.e., has lots of zeros) or conditionally sparse (i.e., has lots of zeros once we condition on some variables like the common risk factors).

To proceed under sparsity, the key job is to find out the location of non-zero entries. There are two broad categories of regularisation, namely banding and shrinkage. Banding is applicable when X_{it} is indexed, and we can expect that larger $|i-j|$ implies lower correlation. Such structure is appropriate for applications where there are natural orderings of variables, such as climatology and spectroscopy. **bickel2008regularized** regularize the $\hat{\Sigma}_X$ in a way that, for a typical entity i, only the set of neighours defined by $\{j \in N_i : |i-j| \leq \rho, j = 1, \ldots, N\}$ have non-zero covariances with i. Other js with $|i-j| > k$ have $\sigma_{ij} = 0$. When X_{it} is not indexed, we may apply some shrinkage techniques to the off-diagonal elements of $\hat{\Sigma}_X$, and achieve sparsity by setting relatively unimportant entries that are smaller than a data-driven threshold to zeros. In general, we have $\tilde{\Sigma}_X = [\tilde{\sigma}_{ij}]_{N\times N}$, where

$$\tilde{\sigma}_{ij} = \begin{cases} \hat{\sigma}_{ii} & i = j \\ s_\lambda(\hat{\sigma}_{ij}) & i \neq j \end{cases}, \tag{2}$$

where $s_\lambda(\cdot)$ is a shrinkage function with λ being a tuning parameter, satisfying (1) $|s_\lambda(\mu)| \leq |\mu|$ for $\mu \in \mathcal{R}$; (2) $s_\lambda(\mu)$ if $|u| \leq \lambda$; (3) $|s_\lambda(\mu) - \mu| \leq \lambda$. **bickel2008covariance** develop a universal threshold for all entries. **cai2011adaptive** establish entry-adaptive threshold to $\hat{\Sigma}_X$. **fan2013large** argue that common factors should be extract first before applying thresholding selection when there are "extremely spiked" eigenvalues in $\hat{\Sigma}_X$ (i.e., the covariance matrix is conditionally sparse).

Another way of shrinkage is via penalization. **yuan2007model**, **d2008first**, **rothman2008sparse** and **friedman2008sparse** among others employ LASSO type penalty to shrink sample covariance matrix. Another seminal penalty choice is SCAD, which is proposed by **fan2001variable**, can be found in **fan2009network** and **lam2009sparsistency**.

However, most of these shrinkage approaches rely on X only and ignore information about pairwise relationships beyond the observed data X, which may suffer a loss especially in a information-abundant era. When X_{it} is not indexed, we are not be able to have strong implications in terms of the locations of non-zeros as in **bickel2008regularized**. However, we do have some idea that who might be connected with whom using augment information apart from X. To proxy for pairwise connectivity among entities, apart from purely statistical methods, there have been other ways. **hoberg2016text** use textual analysis to identify peers. Kaustia and Rantala, 2013 identify peers analyst co-coverage, and **ge2021news** identify peers using business news co-mentioning. Those network information gathered from other sources can help us to identify the locations of non-zeros apart from the information from X itself. Similar location-based thresholding ideas have applied in **fan2016incorporating** and **brownlees2020community**. **fan2016incorporating** apply a hard thresholding method in a way that $\sigma_{ij} = 0$ when i and j are from different sector/industry. **brownlees2020community** first detect community structure using a spectral clustering-based procedure, and then apply a block-by-block thresholding to the off-diagonal elements of $\hat{\Sigma}_X$. In particular, they do not apply and thresholding to $\hat{\sigma}_{ij}$ if i and j are from the same community.

In this paper, we utilize granular network information gathered from other sources that could imply the locations of non-zeros in the de-factored residuals covariance matrix. The first candidate is the new-implied network. It has been documented that common news coverage reveals information about linkages among companies, which are related to many economically important relationships like business alliances, partnerships, banking and financing, customer-supplier, and production similarity (**scherbina2015economic**,

schwenkler2019network). **ge2021news** document that stocks linked by news co-mentioning exhibit additional co-movement beyond what can be explained by common risk factors. Same as **ge2021news**, we use news data from RavenPack Equity files Dow Jones Edition for the period between the beginning of 2004 to the end of 2015. This comprehensive news dataset combines relevant content from multiple sources, including Dow Jones Newswires, Wall Street Journal, and Barron's MarketWatch, which produce the most actively monitored streams of news articles in the financial system. We identify linkages among firms by news co-mentioning. The second candidate network is IBES analyst co-coverage network. Israelsen, 2016 documents that stocks linked by analysts exhibit excess comovement. To construct the analyst co-coverage-based adjacency matrix, we use the Institutional Brokers Estimate System (IBES) detail history files. (after getting the network, our procedure)?

Although here we are applying augment network information to the estimation of large static covariance matrix, similar idea can be extended to the estimation of large dynamic covariance matrix. For example, dynamic network information could be well incorporated into the conditioning information set in **chen2019new**.

3 Data

We consider daily returns of *S&P* 500 stocks for our application. All the stock market related data are from the Center for Research in Security Prices (CRSP). Daily factor returns are obtained from Kenneth French's website.

3.1 News Implied Network

The news data are obtained from RavenPack Equity files Dow Jones Edition for the period January 2004 to December 2015. This comprehensive news dataset combines relevant content from multiple sources, including Dow Jones Newswires, Wall Street Journal, and Barron's MarketWatch, which produce the most actively monitored streams of news articles in the financial system. Each unique news story (identified by a unique story ID) tags the companies mentioned in the news by their unique and permanent entity identifier codes (RP_ENTITY_ID), by which we link to stock identifier TICKER and PERMNO.

As as **ge2021news**, we identify links by news co-mentioning. That is, if a piece of business news reports two companies together, they share a link. We do not consider news that co-mention more than two companies since although news they may carry potential information about links, they provide noisier information. We also remove news with topics including analyst recommendations, rating changes, and index movements as these types of news might stack multiple companies together when they actually do not have real links. **??** provides descriptive statistics for RavenPack Equity files Dow Jones Edition dataset during the sample period. Since our comprehensive news dataset combines several sources, given a similar length of sample period, the number of unique news stories is more than ten times larger than that from **scherbina2015economic** and more than eight hundred times than that from **schwenkler2019network**. For link identification purposes, we only use sample news (1) are not about topics mentioned above (2) tag $S\&P$ 500 companies and (3) mention exactly two companies, which is a subsample of $1,637,256$ unique news stories.

3.2 IBES Analyst Coverage Network

We use the Institutional Brokers Estimate System (IBES) detail history files to construct the analyst co-coverage-based adjacency matrix. For each year in the sample, we consider a stock is covered by an analyst if the analyst issues at least one FY1 or FY2 earnings forecast for the stock during the year. And we consider two stocks as linked if there are common analysts during the year, weighted by the number of common analysts.

4 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:

$$\hat{\sigma}_{ij}^{*} = s_{t_{ij}}(\hat{\sigma}_{ij}) \tag{3}$$

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.

5 Simulations

We demonstrate the network guided estimator and examine its small-sample performance using the following simple simulations. First, we consider the case where the true covariance Σ comes from an AR(1) model. So for $\{(i,j) : i = 1, \ldots, N, j = 1, \ldots, N\}$, $\sigma_{ij}^2 = \sigma_i \sigma_j \rho_{ij}$ and $\rho_{ij} = \rho^{|i-j|}$. We take $N = 400$.

$$S_{ij} = 3 * \rho^{|i-j|}$$

and assume we observe a matrix $G(l)$ indicating the location of highly correlated pairs $L_{ij}(l) = \mathbf{1}\{\rho_{ij} \geq l\}$. Conditional on $L_{ij} = 1$, we observe $G_{ij} = 1$ with probability p and conditional on $L_{ij} = 0$, $G_{ij} = 1$ with probability q. Hence p, q reflect the probability of missing important locations and including false important locations respectively.

We then generate $T = 100$ independent drws of observations $X \sim N(0, \Sigma)$ and estimate Σ using 1. Sample covariance ; 2. Linear Shrinkage estimator; 3. Nonlinear Shrinkage estimator; 4. Universal thresholding on the correlation; 5. and Network Guided thresholding estimator. We now compare their performance. It's worth collecting here the parameters that we will adjust in the experiments:

Parameter	Description
ρ	Determines how strong the correlation is and the sparsity of the covariance matrix Σ
l	Observation level, determines how we classify a pair (i, j) as important, i.e., $L_{ij} = 1$.
p	Conditional on $L_{ij} = 1$, the probability of actually observing $G_{ij} = 1$.
q	Conditional on $L_{ij} = 0$, the probability of observing $G_{ij} = 1$
τ	The Threshold level when we apply generalized thresholding operator on σ_{ij} where $G_{ij} = 0$.

Table 1: Description of varying parameters.

Table 2: The estimation error of various estimators in terms of the Frobenius Norm

rho	Threshold Level	Sample Cov	Linear Shrinkage	Nonlinear Shrinkage	Universal Threshold	Network Guided
0.70	0.0	59.43	42.23	41.61	59.43	59.43
	0.1	59.43	42.23	41.61	49.91	49.85
	0.2	59.43	42.23	41.61	41.94	41.67
	0.3	59.43	42.23	41.61	35.66	34.97
	0.4	59.43	42.23	41.61	31.20	29.81
	0.5	59.43	42.23	41.61	28.56	26.18
	0.6	59.43	42.23	41.61	27.54	23.93
	0.7	59.43	42.23	41.61	27.78	22.81
	0.8	59.43	42.23	41.61	28.83	22.48
	0.9	59.43	42.23	41.61	30.33	22.59
	1.0	59.43	42.23	41.61	32.08	22.98
0.80	0.0	62.54	47.59	46.59	62.54	62.54
	0.1	62.54	47.59	46.59	52.60	52.80
	0.2	62.54	47.59	46.59	44.30	44.52
	0.3	62.54	47.59	46.59	37.89	37.85
	0.4	62.54	47.59	46.59	33.55	32.85
	0.5	62.54	47.59	46.59	31.32	29.48
	0.6	62.54	47.59	46.59	30.95	27.58
	0.7	62.54	47.59	46.59	31.93	26.78
	0.8	62.54	47.59	46.59	33.75	26.71
	0.9	62.54	47.59	46.59	36.02	27.05
	1.0	62.54	47.59	46.59	38.49	27.59
0.90	0.0	63.06	53.18	52.83	63.06	63.06
	0.1	63.06	53.18	52.83	53.98	54.60
	0.2	63.06	53.18	52.83	46.91	47.84

Continued on next page

Table 2: The estimation error of various estimators in terms of the Frobenius Norm

\$rho\$	Threshold Level	Sample Cov	Linear Shrinkage	Nonlinear Shrinkage	Universal Threshold	Network Guided
	0.3	63.06	53.18	52.83	42.16	42.89
	0.4	63.06	53.18	52.83	39.78	39.67
	0.5	63.06	53.18	52.83	39.57	37.96
	0.6	63.06	53.18	52.83	41.08	37.43
	0.7	63.06	53.18	52.83	43.74	37.69
	0.8	63.06	53.18	52.83	47.08	38.41
	0.9	63.06	53.18	52.83	50.74	39.33
	1.0	63.06	53.18	52.83	54.57	40.32
0.95	0.0	57.97	52.58	51.77	57.97	57.97
	0.1	57.97	52.58	51.77	51.42	52.21
	0.2	57.97	52.58	51.77	47.65	48.39
	0.3	57.97	52.58	51.77	46.74	46.40
	0.4	57.97	52.58	51.77	48.35	45.96
	0.5	57.97	52.58	51.77	51.82	46.66
	0.6	57.97	52.58	51.77	56.46	48.04
	0.7	57.97	52.58	51.77	61.73	49.73
	0.8	57.97	52.58	51.77	67.33	51.51
	0.9	57.97	52.58	51.77	73.05	53.24
	1.0	57.97	52.58	51.77	78.77	54.84
0.99	0.0	104.67	115.10	106.32	104.67	104.67
	0.1	104.67	115.10	106.32	114.91	106.31
	0.2	104.67	115.10	106.32	125.35	108.13
	0.3	104.67	115.10	106.32	135.88	110.08
	0.4	104.67	115.10	106.32	146.37	112.10
	0.5	104.67	115.10	106.32	156.67	114.08

Continued on next page

Table 2: The estimation error of various estimators in terms of the Frobenius Norm

\$rho\$	Threshold Level	Sample Cov	Linear Shrinkage	Nonlinear Shrinkage	Universal Threshold	Network Guided
	0.6	104.67	115.10	106.32	166.73	115.97
	0.7	104.67	115.10	106.32	176.56	117.77
	0.8	104.67	115.10	106.32	186.19	119.50
	0.9	104.67	115.10	106.32	195.60	121.16
	1.0	104.67	115.10	106.32	204.74	122.77

Table 3: The estimation error of various estimators in terms of the Matrix-1 Norm

\$rho\$	Threshold Level	Sample Cov	Linear Shrinkage	Nonlinear Shrinkage	Universal Threshold	Network Guided
0.70	0.0	63.77	33.42	33.94	63.77	63.77
	0.1	63.77	33.42	33.94	49.17	49.42
	0.2	63.77	33.42	33.94	37.39	38.23
	0.3	63.77	33.42	33.94	28.37	29.20
	0.4	63.77	33.42	33.94	22.21	22.31
	0.5	63.77	33.42	33.94	18.53	17.86
	0.6	63.77	33.42	33.94	16.18	14.59
	0.7	63.77	33.42	33.94	14.59	12.45
	0.8	63.77	33.42	33.94	13.41	11.03
	0.9	63.77	33.42	33.94	12.65	10.22
	1.0	63.77	33.42	33.94	12.56	9.85
0.80	0.0	73.98	44.02	42.83	73.98	73.98
	0.1	73.98	44.02	42.83	58.63	59.50

Continued on next page

Table 3: The estimation error of various estimators in terms of the Matrix-1 Norm

rho	Threshold Level	Sample Cov	Linear Shrinkage	Nonlinear Shrinkage	Universal Threshold	Network Guided
	0.2	73.98	44.02	42.83	45.86	47.60
	0.3	73.98	44.02	42.83	35.31	37.91
	0.4	73.98	44.02	42.83	26.69	30.04
	0.5	73.98	44.02	42.83	23.68	24.71
	0.6	73.98	44.02	42.83	21.47	20.60
	0.7	73.98	44.02	42.83	20.10	17.40
	0.8	73.98	44.02	42.83	19.66	15.48
	0.9	73.98	44.02	42.83	19.47	14.92
	1.0	73.98	44.02	42.83	19.72	14.85
0.90	0.0	69.65	60.41	60.57	69.65	69.65
	0.1	69.65	60.41	60.57	57.72	58.98
	0.2	69.65	60.41	60.57	50.38	49.66
	0.3	69.65	60.41	60.57	46.30	42.43
	0.4	69.65	60.41	60.57	43.35	38.31
	0.5	69.65	60.41	60.57	41.31	35.10
	0.6	69.65	60.41	60.57	39.91	32.66
	0.7	69.65	60.41	60.57	39.22	31.48
	0.8	69.65	60.41	60.57	39.73	30.93
	0.9	69.65	60.41	60.57	40.38	30.56
	1.0	69.65	60.41	60.57	41.17	30.41
0.95	0.0	95.18	94.47	92.88	95.18	95.18
	0.1	95.18	94.47	92.88	88.39	85.83
	0.2	95.18	94.47	92.88	81.86	76.60
	0.3	95.18	94.47	92.88	76.00	68.02
	0.4	95.18	94.47	92.88	71.17	60.34

Continued on next page

Table 3: The estimation error of various estimators in terms of the Matrix-1 Norm

rho	Threshold Level	Sample Cov	Linear Shrinkage	Nonlinear Shrinkage	Universal Threshold	Network Guided
	0.5	95.18	94.47	92.88	67.30	57.41
	0.6	95.18	94.47	92.88	69.93	56.36
	0.7	95.18	94.47	92.88	73.19	56.04
	0.8	95.18	94.47	92.88	76.37	55.70
	0.9	95.18	94.47	92.88	79.61	55.75
	1.0	95.18	94.47	92.88	82.59	56.93
0.99	0.0	49.73	43.91	50.83	49.73	49.73
	0.1	49.73	43.91	50.83	46.98	48.49
	0.2	49.73	43.91	50.83	58.00	47.60
	0.3	49.73	43.91	50.83	70.41	50.18
	0.4	49.73	43.91	50.83	82.87	54.76
	0.5	49.73	43.91	50.83	94.81	58.76
	0.6	49.73	43.91	50.83	105.96	61.89
	0.7	49.73	43.91	50.83	115.92	63.92
	0.8	49.73	43.91	50.83	128.73	66.98
	0.9	49.73	43.91	50.83	141.60	69.59
	1.0	49.73	43.91	50.83	154.46	71.30

In Table 2, we show the general performance of these estimators when we simulate using different ρ and thresholding level τ. Here we have taken the thresholding operator to be soft thresholding. It can be seen that generally speaking, when the covariance matrix becomes denser, linear, nonlinear shrinakge estimators and the sample covariance estimator become superiro to

Then we consider simulations with varying observation levels l. In Figure 1 when we set observation level equal to 0, the network guided estimator will be the same as the sample covariance estimator, on the other extreme, when observation level is set to 1, the network guided estimator is equivalent to universal thresholding. In between

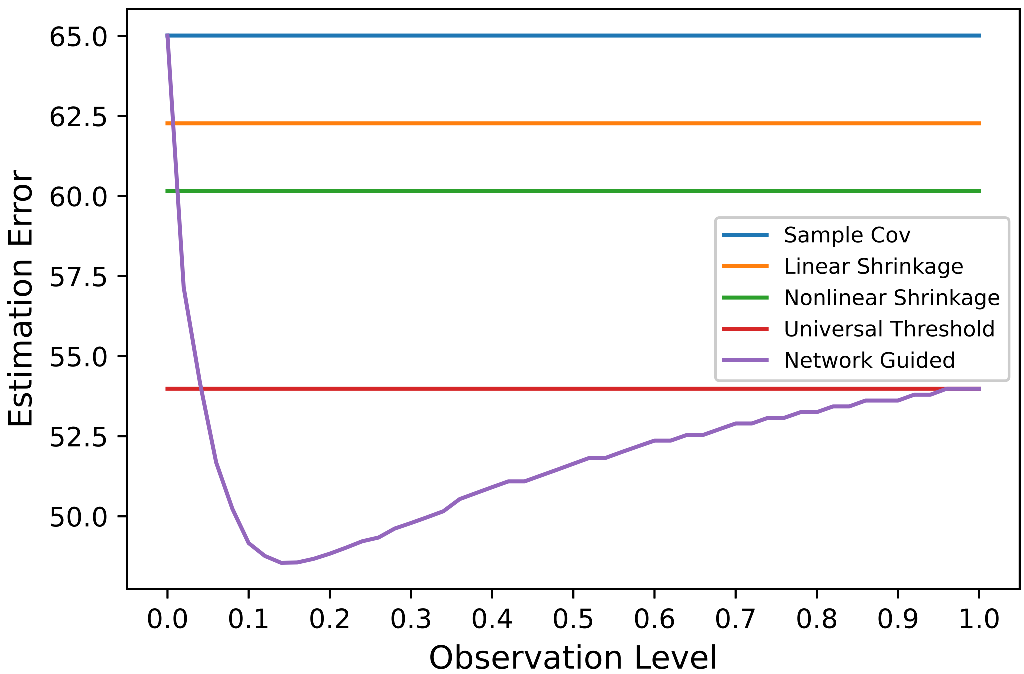

Figure 1: The estimation error against the observation level

these cases, when we have information about the locations of the important pairs, we have a range where the estimation error is lowered.

Table 4: The estimation error of the Network Guided Estimator with varying probabilities p, q that determine how G is generated.

q p	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	47.09	48.44	49.91	51.19	52.43	53.68	54.89	56.14	57.37	58.50	59.61
0.1	46.69	48.00	49.51	50.80	52.11	53.36	54.62	55.92	57.09	58.09	59.32
0.2	46.35	47.81	49.19	50.38	51.73	53.10	54.43	55.47	56.76	57.80	58.99
0.3	45.91	47.40	48.82	50.12	51.50	52.80	53.88	55.10	56.39	57.47	58.79
0.4	45.53	47.08	48.40	49.61	51.01	52.37	53.61	54.72	56.00	57.14	58.45
0.5	45.15	46.74	48.00	49.46	50.71	52.16	53.33	54.60	55.77	57.00	57.99
0.6	44.78	46.14	47.79	48.98	50.38	51.47	52.93	54.10	55.32	56.59	57.85
0.7	44.37	45.94	47.10	48.96	50.10	51.29	52.52	53.96	55.03	56.23	57.44
0.8	44.06	45.58	46.85	48.36	49.66	50.87	52.15	53.62	54.81	55.90	57.22
0.9	43.51	45.06	46.28	48.02	49.33	50.56	51.99	53.27	54.46	55.69	56.82
1.0	43.14	44.71	46.08	47.52	48.87	50.15	51.69	52.80	54.13	55.29	56.53

In Table 4, we have when $p = q = 0$ the estimation error of the universal thresholding estimator, and $p = q = 1$ the sample covaraince estimation error. As we can see, as long as q is not large, the estimation error will be smaller when we have a higher probability p of observing the true large elements. It should be noted that q in fact cannot be very large, given that the whole matrix is sparse.

6 Empirical Studies

6.1 Global Minimum Variance Portfolio

We apply the Network Guided Estimator to a portfolio management similar to Ledoit and Wolf, 2004. We collect daily return data on SP500 stock from 2004 to 2019 from CRSP, together with daily data on Fama-French 3 factors and the risk free rate.

Assume that the excess returns follow the following factor model

$$Y_{it} = B_i' F_t + \varepsilon_{it}$$

and we assume that $\Sigma = [E\varepsilon_i\varepsilon_j]_{1\le i,j\le N}$ is sparse.

We do a rolling window analysis, each window consists of an estimation period of 252 days and a testing period of 21 days. In the estimation period, we estimate the factor loadings by linear time series regression of excess return Y_{it} on F_t, hence allowing the betas to vary over time, and find the de-factored excess return by

$$\hat{\varepsilon}_{it} = Y_{it} - \hat{B}_i' F_t$$

and in order to estimate the covariance matrix of $Y = (Y_1, \ldots, Y_N)'$, we have, under the assumption that ε's are independent of F_t,

$$\Sigma_Y = B\Sigma_F B' + \Sigma_\varepsilon$$

We replace the factor covariance component by $\hat{B}\hat{\Sigma}_F\hat{B}$, where $\hat{\Sigma}_F$ is the sample covariance of factors in that period, and we estimate Σ_ε by the Network Guided Estimator applied to $\hat{\Sigma}_\varepsilon = \frac{1}{T}\sum_t \hat{\varepsilon}_t\hat{\varepsilon}_t'$.

In order to apply the Network Guided Estimator, we consider two G matrix that comes from analysts co-coverage:1. IBES: .... 2: Dow Jones: ....

In order to keep G sparse and mitigate the noisy observations, we set for IBES, $G_{ij} = 1$ if firms (i, j) are mentioned more than 18 times, and for Dow Jones data, $G_{ij} = 1$ if firms (i,j) are co-mentioned more than 100 times. With this choice, the total number of links is around 1‰of the whole network matrix. We select the thresholding parameter using cross-validation with the constraint that the resulting estimate is positvie definite. When the thresholding level becomes higher, the resulting estimate becomes more sparse, in the limit, it'll be a diagonla matrix and p.d., so we keep the thresholding level above the minimum level for the estimate to be p.d.

After using the data from estimation period to estimate $\hat{B}, \hat{\Sigma}_F, \hat{\Sigma}_\varepsilon$, we construct

$$\hat{\Sigma}_Y = \hat{B}\hat{\Sigma}_F\hat{B}' + \hat{\Sigma}_\varepsilon$$

and construct the *global minimum variance* portfolio where weights are given by

$$w = \frac{\hat{\Sigma}_Y \mathbf{1}}{\mathbf{1}' \hat{\Sigma}_Y \mathbf{1}}$$

where $\mathbf{1}$ is a conforming vector of ones. We collect the portfolio return over the next 21-day testing period. This conludes one of the rolling windows. Then we move forward 21 days and repeat this exercise. Using 2004- 2019 daily data, we can construct a daily portfolio return from 2005 to 2019, where the portfolio is rebalanced every 21 days. We compute the holding period return of this portfolio and its standard deviation. In autoreft:4 we show the result together with mean and Sharpe ratio and compare it with global minimum variance portfolio constructed using linear shrinkage and universal thresholding. It's worth mentioning that given we are comparing global minimum variance portfolio, the standard deviation is the relevant indicator of performance.

7 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, 7).

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, 8).

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. 8).

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. 8).

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. 8).

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 p. 8).

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. 8).

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. 8).

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

Israelsen, Ryan D (2016). "Does common analyst coverage explain excess comovement?" In: *Journal of Financial and Quantitative Analysis*, pp. 1193–1229 (cit. on pp. 2, 6).

Kaustia, Markku and Ville Rantala (2013). "Common analyst-based method for defining peer firms". In: *Available at SSRN* (cit. on p. 5).

— (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. 8).

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. 8, 16, 18).

— (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. 8).

— (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. 8, 18).

Pourahmadi, Mohsen (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data*. John Wiley & Sons (cit. on p. 7).

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, 8).

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. 8).