Sparse Approximate Factor Estimation for High-Dimensional Covariance Matrices

Maurizio Daniele Winfried Pohlmeier Aygul Zagidullina

University of Konstanz

4th Konstanz - Lancaster Workshop on Finance and Econometrics

July 30th, 2018

Outline

Introduction

- In the recent years, the estimation of high-dimensional covariance matrices and their inverses (precision matrices) has received a great attention
- In economics and finance it is central for portfolio allocation, risk measurement, asset pricing and graphical network analysis
- The list of important applications from other areas of research includes, for example, the analysis of climate data, gene classification and image classification

Introduction

In the high dimensional setting ($T \ll N$) the sample covariance matrix has a number of undesirable properties:

- the sample covariance matrix becomes nearly singular and estimates the population covariance matrix poorly

- even if it is invertible, it is ill-conditioned, meaning that the inverse amplifies the estimation error

Methodologies that mitigate these shortcomings can be classified into the following groups:

- **Covariance Shrinkage Methods**: The shrinkage estimator is a linear combination of the sample covariance matrix and another estimator (Ledoit and Wolf (2003), Kourtis, Dotsis, and Markellos (2012))
- **Sparse Covariance Matrix Estimators**: The off-diagonal elements of the covariance matrix are shrunken to zero (Rothman, Levina, and Zhu (2009), Bien and Tibshirani (2011))
- **Factor Models**: The covariance matrices are implied by the large dimensional factor models, either with observable or latent factors (Sharpe (1963), Fama and French (1993), Chamberlain and Rothschild (1983), Fan, Liao, and Mincheva (2013))

Introduction - Approximate Factor Model

The approximate factor model is defined as:

$$X = \Lambda F' + u,$$

where

- Λ is a $N \times r$ matrix of unobserved factor loadings
- F is a $T \times r$ matrix of unobserved factors
- $r \ll N$ is the number of underlying factors
- u is a $N \times T$ matrix of idiosyncratic errors that may incorporate some serial and cross-sectional correlations

Introduction - Strong and weak factors

- In order to achieve consistent parameter estimates using PCA, the crucial assumption of pervasive (strong) factors is necessary (Onatski (2012))

$$\pi_{\max}\left(\Lambda'\Lambda\right) = \mathcal{O}\left(N^{\beta}\right), \quad \text{for } 0 \leq \beta \leq 1,$$

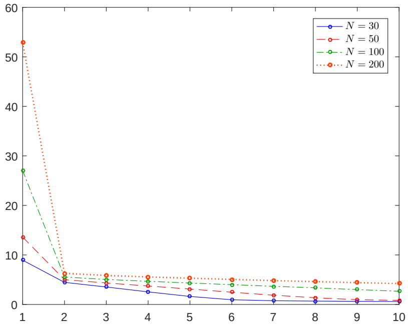
  where $\beta = 1$ denotes the strong factor case. $0 \leq \beta < 1$ specifies the weak factor framework, where factors may only explain a subset of the analyzed dataset.

- Implied sparsity for Λ

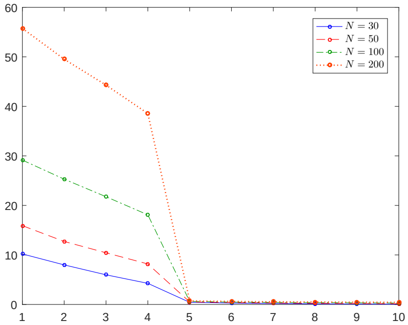
$$\max_{k \leq r} \sum_{i=1}^{N} |\lambda_{ik}| \leq \sqrt{N \pi_{\max}\left(\Lambda'\Lambda\right)} = \mathcal{O}\left(N^{(1+\beta)/2}\right)$$

- However, the strong factor assumption is hardly observed for real datasets

Introduction - Eigenvalue Distribution for strong factors



(a) Eigenvalues for simulated data with 1 strong factor with $T = 450$



(b) Eigenvalues for simulated data with 4 strong factors with $T = 450$

Introduction - Eigenvalue comparison

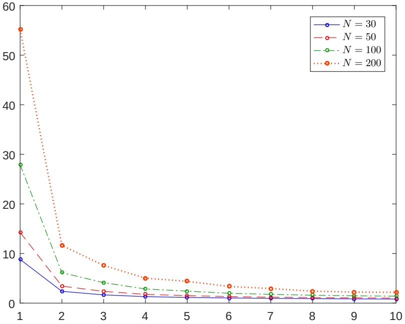


Figure: Eigenvalues for stock returns based on the S&P 500 index with $T = 450$

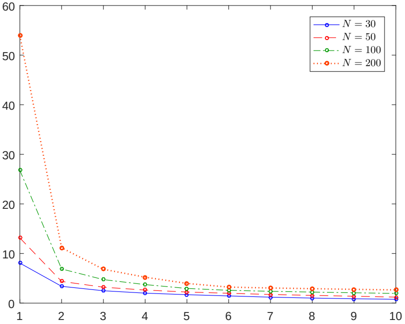


Figure: Eigenvalues for simulated data with 1 strong factor and 3 weak factors with $T = 450$

Introduction - Sparse Factor Model

- Introduce a sparse factor model that allows for sparsity in the factor loadings matrix
- The model specification allows for a weak factor framework, where not all factors explain the entire set of N variables
- On the contrary, some factors may explain only the variation for specific sectors (i.e. Financial or Industrial sector)
- We show consistency for the sparse factor model estimator
- The corresponding covariance matrix estimator is also consistent

Model Specification - Log-Likelihood function

Negative Log-Likelihood for the data covariance matrix

$$\mathcal{L}(\Lambda, \Sigma_F, \Sigma_u) = \frac{1}{N} \log \left| \det \left(\Lambda \Sigma_F \Lambda' + \Sigma_u \right) \right| + \frac{1}{N} \text{tr} \left[S_x \left(\Lambda \Sigma_F \Lambda' + \Sigma_u \right)^{-1} \right]$$

where

- Σ_F and Σ_u are the covariance matrices of the factors and the idiosyncratic component, respectively
- S_x denotes the sample covariance matrix of the data

Identifying restrictions:

$$\Sigma_F = I_r \quad \text{and} \quad \Lambda' \Sigma_u^{-1} \Lambda \ \text{ is diagonal}$$

Model Specification - Sparse Factor Model

To allow for sparsity in Λ we consider the following L_1-penalized maximum likelihood problem:

$$\mathcal{L}(\Lambda, \Phi_u) = \frac{1}{N} \log \left| \det \left(\Lambda\Lambda' + \Phi_u \right) \right| + \frac{1}{N} \text{tr} \left[S_x \left(\Lambda\Lambda' + \Phi_u \right)^{-1} \right] + \frac{\mu}{N} \sum_{k=1}^{r} \sum_{i=1}^{N} \left| \lambda_{ik} \right|,$$

where

- $\Phi_u = \text{diag}\left(\Sigma_u\right)$
- μ is a regularization parameter
- λ_{ik} is the i^{th} factor loading in the k^{th} column

Model Specification - Estimation of Σ_u

To allow for cross-sectional correlation in idiosyncratic error covariance matrix, we use the principal orthogonal complement thresholding (POET) estimator by Fan et al. (2013). More specifically, the estimated idiosyncratic error covariance matrix $\hat{\Sigma}_u^{\tau}$ based on the POET method is defined as:

$$\hat{\Sigma}_u^{\tau} = \hat{\sigma}_{ij}^{\tau}, \quad \hat{\sigma}_{ij}^{\tau} = \begin{cases} \hat{\sigma}_{u,ii}, & i = j \\ \mathcal{S}(\hat{\sigma}_{u,ij}, \tau), & i \neq j \end{cases}$$

where $\hat{\sigma}_{u,ij}$ is the ij-th element of the sample covariance matrix S_u of the estimated factor model residuals, $\tau = \frac{1}{\sqrt{N}} + \sqrt{\frac{\log N}{T}}$ and $\mathcal{S}$ denotes the soft-thresholding operator:

$$\mathcal{S}(\sigma_{u,ij}, \tau) = \text{sign}(\sigma_{u,ij})(|\sigma_{u,ij}| - \tau)_+.$$

Sparse Factor Model based covariance matrix estimator

The covariance matrix of the dataset X according to the approximate factor model is:

$$\Sigma = \Lambda \Sigma_F \Lambda' + \Sigma_u$$

Inserting the estimates of the sparse factor model yields:

$$\hat{\Sigma}_{SAF} = \hat{\Lambda} \hat{\Sigma}_{\hat{F}} \hat{\Lambda}' + \hat{\Sigma}_u^{\tau},$$

- $\hat{\Sigma}_{\hat{F}}$ is sample covariance matrix of the estimated factors
- $\hat{\Sigma}_u^{\tau}$ is obtained by the thresholding estimator based on the residuals of the sparse factor model

Theoretical Properties - Assumption (1)

Assumption (Weakness of Factor Loadings)

There exists a constant $c > 0$ *such that for all* N,

$$c^{-1} < \pi_{\min}\left(\frac{\Lambda'\Lambda}{N^{\beta}}\right) \leq \pi_{\max}\left(\frac{\Lambda'\Lambda}{N^{\beta}}\right) < c,$$

where $1/2 \leq \beta \leq 1$.[a]

[a] The lower limit 1/2 for β is necessary to consistently estimate the factors.

Theoretical Properties - Assumption (2)

Assumption (Data generating process)

(i) $\{u_t, f_t\}_{t\geq 1}$ *is strictly stationary. In addition,* $\mathbf{E}[u_{it}] = \mathbf{E}[u_{it} f_{kt}] = 0$, *for all* $i \leq N$, $k \leq r$ *and* $t \leq T$.

(ii) *There exists* $r_1, r_2 > 0$ *and* $b_1, b_2 > 0$, *such that for any* $s > 0$, $i \leq N$ *and* $k \leq r$,

$$\mathbf{P}(|u_{it}| > s) \leq \exp(-(s/b_1)^{r_1}), \quad \mathbf{P}(|f_{kt}| > s) \leq \exp(-(s/b_2)^{r_2})$$

(iii) *Strong mixing*

(iv) *There exist constants* $c_1, c_2 > 0$ *such that* $c_2 \leq \pi_{\min}(\Sigma_{u0}) \leq \pi_{\max}(\Sigma_{u0}) \leq c_1$.

Theoretical Properties - Assumption (3)

Assumption (Sparsity)

(i) $L_N = \sum_{i=1}^{N} \mathbb{1}\{\lambda_{ik} \neq 0\} = \mathcal{O}(N), \quad \forall k = 1, \ldots, r$

(ii) $S_N = \max_{i \leq N} \sum_{j=1}^{N} \mathbb{1}\{\sigma_{u,ij} \neq 0\}$,

where $\mathbb{1}\{\cdot\}$ defines an indicator function that is equal to one if the boolean argument in braces is true.

Theoretical Properties - Convergence rates (1)

Consistency of the Sparse Approximate Factor Model Estimator:

Theorem (1)

Under Assumptions 1, 2 and 3 the sparse factor model satisfies for T and $N \to \infty$, the following properties:

$$\frac{1}{N}\left\|\hat{\Lambda} - \Lambda_0\right\|_F^2 = \mathcal{O}_p\left(\mu^2 + \frac{\log N^\beta}{N} + \frac{1}{N^\beta}\frac{\log N}{T}\right)$$

and

$$\frac{1}{N}\left\|\hat{\Phi}_u - \Phi_{u0}\right\|_F^2 = \mathcal{O}_p\left(\frac{\log N^\beta}{N} + \frac{\log N}{T}\right),$$

for $1/2 \leq \beta \leq 1$.

Theoretical Properties - Convergence rates (2)

Consistency of the Sparse Approximate Factor Model Estimator:

Theorem (1)

Hence, for $\log(N) = o(T)$ *and the regularization parameter* $\mu = o(1)$*, we have:*

$$\frac{1}{N}\left\|\hat{\Lambda} - \Lambda_0\right\|_F^2 = o_p(1), \quad \frac{1}{N}\left\|\hat{\Phi}_u - \Phi_{u0}\right\|_F^2 = o_p(1), \quad \left\|\hat{f}_t - f_{t0}\right\| = o_p(1)$$

For the idiosyncratic error covariance matrix estimator in the second step, we get:

$$\left\|\hat{\Sigma}_u^{\tau} - \Sigma_{u0}\right\| = \mathcal{O}_p\left(S_N\sqrt{\mu^2 + \frac{N}{L_N}d_T}\right),$$

where $d_T = \frac{\log N^{\beta}}{N} + \frac{1}{N^{\beta}}\frac{\log N}{T}$.

Theoretical Properties - Convergence rates (3)

Convergence Rates for the Covariance Matrix Estimator:

Theorem (2)

Under Assumptions 1, 2 and 3, the covariance matrix estimator based on the sparse factor model satisfies for T, $N \to \infty$ and $1/2 \leq \beta \leq 1$, the following properties:

$$\frac{1}{N}\left\|\hat{\Sigma}_{\mathrm{SAF}}-\Sigma\right\|_{\Sigma}^{2}=\mathcal{O}_p\left(\left[\mu^2+d_T\right]^2+\left[\frac{N^{\beta}}{N}+\frac{S_N^2}{N}\right]\left[\mu^2+d_T\right]\right)$$

and

$$\frac{1}{N}\left\|\hat{\Sigma}_{\mathrm{SAF}}-\Sigma\right\|_{F}^{2}=\mathcal{O}_p\left(N\left[\mu^2+d_T\right]^2+\left[N^{\beta}+S_N^2\right]\left[\mu^2+d_T\right]\right),$$

where $d_T=\frac{\log N^{\beta}}{N}+\frac{1}{N^{\beta}}\frac{\log N}{T}$ and $\|A\|_{\Sigma}=\frac{1}{\sqrt{N}}\left\|\Sigma^{-1/2}A\Sigma^{-1/2}\right\|_F$ denotes the weighted quadratic norm introduced by Fan, Fan, and Lv (2008).

Theoretical Properties - Convergence rates (4)

Convergence Rates for the Covariance Matrix Estimator:

$$\frac{1}{N}\left\|\hat{\Sigma}_{\mathsf{SAF}}-\Sigma\right\|_F^2=\mathcal{O}_p\left(N\left[\mu^2+d_T\right]^2+N^{\beta}\mu^2+S_N^2\left[\mu^2+d_T\right]\right.$$
$$\left.+\frac{N^{\beta}\log N^{\beta}}{N}+\frac{\log N}{T}\right)$$

where $d_T=\frac{\log N^{\beta}}{N}+\frac{1}{N^{\beta}}\frac{\log N}{T}$.

Implementation

- The optimization is not straightforward as the objective function contains the concave component $\log\left|\det\left(\Lambda\Lambda' + \Sigma_u\right)\right|$ and the convex part $\operatorname{tr}\left[S_x\left(\Lambda\Lambda' + \Sigma_u\right)^{-1}\right]$
- It is possible to solve the problem approximately, by substituting the concave part with its tangent plane:

$$\log\left|\det\left(\hat{\Lambda}_m\hat{\Lambda}'_m + \hat{\Sigma}_{u,m}\right)\right| + \operatorname{tr}\left[2\hat{\Lambda}'_m\left(\hat{\Lambda}_m\hat{\Lambda}'_m + \hat{\Sigma}_{u,m}\right)^{-1}\left(\Lambda - \hat{\Lambda}_m\right)\right],$$

where m denotes the m^{th} iteration step.

Implementation - Majorize-Minimize algorithm

The model is estimated using the Majorize-Minimize algorithm proposed by Bien and Tibshirani (2011).
The majorized likelihood using the tangent plane of $\log\left|\det\left(\Lambda\Lambda' + \Phi_u\right)\right|$ is given by:

$$\bar{\mathcal{L}} = \frac{1}{N}\log\left|\det\left(\hat{\Lambda}_m\hat{\Lambda}_m' + \hat{\Phi}_{u,m}\right)\right| + \frac{1}{N}\mathrm{tr}\left[2\hat{\Lambda}_m'\left(\hat{\Lambda}_m\hat{\Lambda}_m' + \hat{\Phi}_{u,m}\right)^{-1}\left(\Lambda - \hat{\Lambda}_m\right)\right]$$
$$+ \frac{1}{N}\mathrm{tr}\left[S_x\left(\Lambda\Lambda' + \hat{\Phi}_{u,m}\right)^{-1}\right] + \frac{\mu}{N}\sum_{k=1}^{r}\sum_{i=1}^{N}|\lambda_{ik}|$$

Simulation study - Covariance matrix estimation

- In the Monte Carlo simulation we analyze the accuracy of the different methods in estimating the true covariance matrix and consider two different designs for the true covariance matrix:
- **Uniform design**

$$\sigma^{\mathrm{u}}_{ii} = 1 \text{ and } \sigma^{\mathrm{u}}_{ij} = \eta\, \mathcal{U}_{(0,1)}, \quad \text{for } i \neq j,$$

  where $\mathcal{U}_{(0,1)}$ denotes a standard uniform random variable and we set $\eta \in \{0.0, 0.025, 0.05, 0.075\}$.
- **Sparse Covariance Matrix Design**
  The ijth element of the covariance matrix $\sigma_{ij} = \sigma_{ji}$ is assigned to be non-zero with a probability of p, where $p \in \{0, 0.05, 0.1, 0.2\}$. Similar as in the uniform design the diagonal elements are set to 1. The non-zero off-diagonal elements are independently drawn from a uniform distribution $\left(\mathcal{U}_{(0,0.25)}\right)$.

Simulation study - Covariance matrix estimation

- For both covariance designs we draw the random data series from a multivariate normal distribution.
- $N \in \{30, 50, 100, 200\}$
- T is set to 60
- The Frobenius norm is used as evaluation criterion

Simulation study - Results

Table: Results for the uniform covariance matrix design

N	Model	η = 0	η = 0.025	η = 0.05	η = 0.075
30	Sample	15.82	15.72	15.93	15.75
	AFM	6.53	6.66	6.83	6.90
	DFM	6.27	6.42	6.63	6.71
	SAF	**0.45**	**0.62**	**1.09**	**1.95**
	ST	1.02	1.20	1.77	2.86
	BT	2.99	3.18	3.71	4.78
	LW	3.15	3.36	3.90	4.38
	ADZ	2.02	2.24	2.77	3.54
50	Sample	42.96	43.31	42.79	43.32
	AFM	10.73	11.24	11.81	12.41
	DFM	10.60	11.04	11.60	12.26
	SAF	**0.50**	**1.00**	**2.37**	**4.48**
	ST	1.72	2.23	3.69	6.30
	BT	4.99	5.54	7.03	9.67
	LW	4.90	5.94	7.14	8.68
	ADZ	3.33	3.93	5.36	7.33

N	Model	η = 0	η = 0.025	η = 0.05	η = 0.075
100	Sample	1.72E+02	1.71E+02	1.72E+02	1.72E+02
	AFM	23.99	25.91	27.97	32.62
	DFM	23.82	25.61	27.60	32.17
	SAF	**0.64**	**2.65**	**8.16**	**14.88**
	ST	3.36	5.46	11.56	21.88
	BT	10.25	12.31	18.58	28.83
	LW	10.68	15.17	19.98	26.64
	ADZ	6.82	8.75	14.34	21.02
200	Sample	6.81E+02	6.83E+02	6.84E+02	6.84E+02
	AFM	58.44	65.56	77.33	101.78
	DFM	58.03	64.66	76.90	100.87
	SAF	**0.81**	**8.92**	**28.73**	**53.48**
	ST	6.74	15.09	39.93	81.63
	BT	21.26	29.56	54.49	96.56
	LW	26.44	44.57	65.88	92.75
	ADZ	13.89	21.86	42.03	67.39

The sparse factor model (SAF) is compared to the approximate factor model (AFM), the dynamic factor model (DFM), the soft-thresholding estimator (ST) of Rothman et al. (2009), the sparse covariance estimator by Bien and Tibshirani (2011) (BT), the shrinkage estimator by Ledoit and Wolf (2003) (LW), and the design-free estimator by Abadir, Distaso, and Žikeš (2014) (ADZ)

Simulation study - Results

Table: Results for the sparse covariance matrix design

N	Model	p = 0	p = 0.050	p = 0.10	p = 0.200
30	Sample	15.82	15.89	16.14	15.86
	AFM	6.53	11.01	11.45	12.61
	DFM	6.27	11.38	11.18	12.29
	SAF	**0.45**	**3.63**	**4.77**	**6.68**
	ST	1.02	5.60	6.45	8.97
	BT	2.99	7.44	8.27	9.83
	LW	3.15	6.86	6.84	7.39
	ADZ	2.02	6.49	6.45	7.21
50	Sample	42.96	43.62	42.98	43.53
	AFM	10.73	19.42	24.08	25.26
	DFM	10.60	19.76	24.47	25.22
	SAF	**0.50**	**8.21**	**13.17**	**14.85**
	ST	1.72	10.48	15.49	17.06
	BT	4.99	13.81	18.93	20.43
	LW	4.90	12.54	15.43	16.21
	ADZ	3.33	11.44	15.00	15.67

N	Model	p = 0	p = 0.050	p = 0.10	p = 0.200
100	Sample	1.72E+02	1.72E+02	1.71E+02	1.72E+02
	AFM	23.99	45.50	47.07	52.28
	DFM	23.82	46.81	47.05	52.08
	SAF	**0.64**	**20.57**	**22.14**	**28.49**
	ST	3.36	24.45	25.46	32.05
	BT	10.25	31.55	32.52	39.13
	LW	10.68	31.09	31.70	37.51
	ADZ	6.82	27.13	27.53	32.92
200	Sample	6.81E+02	6.84E+02	6.84E+02	6.84E+02
	AFM	58.44	111.01	110.47	113.82
	DFM	58.03	112.33	110.58	113.26
	SAF	**0.81**	**48.93**	**48.57**	**52.89**
	ST	6.74	55.54	55.07	59.45
	BT	21.26	70.50	69.88	74.28
	LW	26.44	78.12	78.72	85.55
	ADZ	13.83	61.36	60.73	64.37

The sparse factor model (SAF) is compared to the approximate factor model (AFM), the dynamic factor model (DFM), the soft-thresholding estimator (ST) of Rothman et al. (2009), the sparse covariance estimator by Bien and Tibshirani (2011) (BT), the shrinkage estimator by Ledoit and Wolf (2003) (LW), and the design-free estimator by Abadir et al. (2014) (ADZ)

Dataset	N	Time Period	# obs.
S&P 500 index	30, 50, 100, 150, 200	01/1974 - 04/2015	496

- Out-of-sample rolling-window forecasting experiment based on GMV portfolios
- The estimation window size is $T = 60$ months
- The stocks for each subset of size N are selected randomly at the beginning of each simulation and kept fixed for the entire forecasting experiment

Empirical Application - Results

Model	1/N	GMVP	**SAF**	AFM	DFM	SFM	FF3F	LW	KDM	AZD
					N = 30					
SD	0.0477	0.0665	**0.0458**	0.0498	0.0484	0.0478	0.0469	0.0489	0.0483	0.0474
AV	0.0084	0.0075	**0.0086**	0.0074	0.0075	0.0085	0.0081	0.0077	0.0081	0.0079
CE	0.0062	0.0031	**0.0065**	0.0049	0.0052	0.0062	0.0059	0.0053	0.0058	0.0056
SR	0.1766	0.1123	**0.1868**	0.1475	0.1553	0.1768	0.1727	0.1581	0.1678	0.1658
					N = 50					
SD	0.0467	0.1165	**0.0446**	0.0496	0.0479	0.0468	0.0458	0.0486	0.0479	0.0470
AV	0.0084	0.0082	**0.0088**	0.0075	0.0075	0.0084	0.0079	0.0080	0.0081	0.0081
CE	0.0062	-0.0055	**0.0068**	0.0050	0.0052	0.0062	0.0058	0.0057	0.0058	0.0059
SR	0.1797	0.0706	**0.1973**	0.1510	0.1564	0.1799	0.1736	0.1651	0.1687	0.1717

Note: SD, AV, CE and SR denote the standard deviation, average returns, certainty equivalent and Sharpe ratio, respectively

Empirical Application - Results

Model	1/N	GMVP	**SAF**	AFM	DFM	SFM	FF3F	LW	KDM	AZD
					N = 100					
SD	0.0462	-	**0.0435**	0.0479	0.0468	0.0463	0.0448	0.0466	0.0463	0.0456
AV	0.0084	-	**0.0092**	0.0078	0.0070	0.0084	0.0078	0.0082	0.0081	0.0081
CE	0.0063	-	**0.0073**	0.0055	0.0048	0.0063	0.0058	0.0060	0.0059	0.0061
SR	0.1825	-	**0.2112**	0.1627	0.1500	0.1827	0.1740	0.1753	0.1740	0.1785
					N = 150					
SD	0.0460	-	**0.0429**	0.0466	0.0459	0.0460	0.0444	0.0453	0.0452	0.0445
AV	0.0084	-	**0.0094**	0.0082	0.0066	0.0084	0.0077	0.0082	0.0080	0.0083
CE	0.0062	-	**0.0075**	0.0060	0.0045	0.0063	0.0057	0.0062	0.0060	0.0063
SR	0.1817	-	**0.2182**	0.1747	0.1438	0.1819	0.1730	0.1811	0.1774	0.1864
					N = 200					
SD	0.0459	-	**0.0426**	0.0459	0.0454	0.0459	0.0440	0.0444	0.0450	0.0440
AV	0.0084	-	**0.0095**	0.0086	0.0063	0.0084	0.0076	0.0083	0.0078	0.0083
CE	0.0063	-	**0.0077**	0.0065	0.0043	0.0063	0.0057	0.0063	0.0058	0.0063
SR	0.1822	-	**0.2238**	0.1883	0.1394	0.1824	0.1728	0.1864	0.1732	0.1883

Note: SD, AV, CE and SR denote the standard deviation, average returns, certainty equivalent and Sharpe ratio, respectively

Conclusion

- We introduce a sparse factor model that allows for sparsity in the factor loadings matrix
- We show consistency for the sparse factor model estimator and corresponding covariance matrix estimator
- The simulation study indicates that our covariance estimator offers the lowest estimation errors in term of the Frobenius norm for different true covariance matrix designs
- It stabilizes the portfolio weights and outperforms in the out-of-sample portfolio application, methods often used in the finance literature in terms of SD, CE and SR.

References I

Karim M Abadir, Walter Distaso, and Filip Žikeš. Design-free estimation of variance matrices. *Journal of Econometrics*, 181(2):165–180, 2014.

Jushan Bai and Kunpeng Li. Statistical analysis of factor models of high dimension. *The Annals of Statistics*, pages 436–465, 2012.

Jacob Bien and Robert J Tibshirani. Sparse estimation of a covariance matrix. *Biometrika*, 98(4):807, 2011.

Gary Chamberlain and Michael Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51 (5):1281–304, 1983.

Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1): 3–56, 1993.

References II

Jianqing Fan, Yingying Fan, and Jinchi Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147 (1):186–197, 2008. ISSN 03044076. doi: 10.1016/j.jeconom.2008.09.017.

Jianqing Fan, Yuan Liao, and Martina Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75 (4):603–680, 2013.

Apostolos Kourtis, George Dotsis, and Raphael N Markellos. Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. *Journal of Banking & Finance*, 36(9):2522–2531, 2012.

Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621, 2003.

References III

Alexei Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258, 2012.

Adam J Rothman, Elizaveta Levina, and Ji Zhu. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186, 2009.

William F Sharpe. A simplified model for portfolio analysis. *Management science*, 9(2):277–293, 1963.

Strong Mixing

Define the mixing coefficient:

$$\alpha(T) = \sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{T}^{\infty}} \left| \mathbf{P}(A)\,\mathbf{P}(B) - \mathbf{P}(AB) \right|,$$

where $\mathcal{F}_{-\infty}^{0}$ and F_{T}^{∞} denote the σ-algebras generated by $\{(f_t, u_t) : -\infty \leq t \leq 0\}$ and $\{(f_t, u_t) : T \leq t \leq \infty\}$
Strong mixing: There exist $r_3 > 0$ and $C > 0$ satisfying: for all $T \in \mathbb{Z}^{+}$,

$$\alpha(T) \leq \exp(-CT^{r_3})$$