Nonlinear shrinkage estimation of large integrated covariance matrices

Nonlinear Shrinkage Estimation of Large Integrated Covariance Matrices

Clifford Lam*, Phoenix Feng†, and Charlie Hu‡

Department of Statistics, London School of Economics and Political Science

Abstract

Integrated covariance matrices arise in intra-day models of asset returns, which allow volatility to change across the trading day. When the number of assets is large, the natural estimator of such a matrix suffers from bias, contributed from extreme eigenvalues. We introduce a novel nonlinear shrinkage estimator for the integrated covariance matrix which shrinks the extreme eigenvalues of a realized covariance matrix back to an acceptable level, and enjoys a certain asymptotic efficiency when the number of assets is of the same order as the number of data points. Novel maximum exposure and actual risk bounds are derived when our estimator is used in constructing the minimum variance portfolio. Compared to other methods, our estimator performs favorably in both simulations and a real data analysis.

Key words and phrases. Extreme eigenvalue; High dimension; Intra-day volatility; Maximum exposure bound; Portfolio allocation; Realized covariance.

*Clifford Lam is Associate Professor, Department of Statistics, London School of Economics. Email: C.Lam2@lse.ac.uk
†Phoenix Feng is PhD student, Department of Statistics, London School of Economics. Email: H.Feng2@lse.ac.uk
‡Charlie Hu is PhD student, Department of Statistics, London School of Economics. Email: Q.Hu2@lse.ac.uk

1 Introduction

Intra-day data on financial asset returns are of increasing interest for portfolio allocation and risk management (Fan et al., 2012). Models for such data need to account for rapid changes in volatility during a trading day. To capture such changes, it is natural to consider covolatility processes and to combine covariances between pairs of asset returns over time through what is called an integrated covariance matrix, defined formally in the next section. There are various challenges to estimating this matrix (Aït-Sahalia et al., 2005, Asparouhova et al., 2013). In this paper, we consider the bias arising when p, the number of assets observed, is large. Specifically, we suppose that p is the same order as the sample size n, i.e., $p/n \to c > 0$ for some constant $c > 0$. If prices are observed at synchronous time points, a natural estimator of the integrated covariance matrix can be obtained from an empirical covariance matrix of the observed returns. However, this estimator suffers from bias, which may be expressed via bias of its extreme eigenvalues (Bai & Silverstein, 2010).

To rectify this bias problem, many researchers have focused on regularized estimation of covariance or precision matrices with special structures, from banded (Bickel & Levina, 2008b) or sparse covariance matrices (Bickel & Levina, 2008a, Cai & Zhou, 2012, Lam & Fan, 2009, Rothman et al., 2008), sparse precision matrices (Friedman et al., 2008, Meinshausen & Bühlmann, 2006), sparse modified Cholesky factors (Pourahmadi, 2007), to a spiked covariance matrix from a factor model (Fan et al., 2008, 2011), or combinations of these (Fan et al., 2013).

Ledoit & Wolf (2012) proposed a nonlinear shrinkage formula for shrinking the extreme eigenvalues in a sample covariance matrix without a particular structure for the true covariance matrix. However, their approach is not applicable to non-identically distributed random vectors, as arise with intra-day return data.

Lam (2016) proves that by splitting the data into two independent portions, one can achieve the same nonlinear shrinkage asymptotically without the need to evaluate a shrinkage formula. In this paper, we modify this method to achieve nonlinear shrinkage of eigenvalues in a covariance matrix. Our method produces a positive definite estimator of the integrated covariance matrix asymptotically almost surely, and involves only eigen-decompositions of matrices of size $p \times p$, which is not computationally expensive when p is of order of hundreds, the typical order for p in portfolio allocation. We also present the maximum exposure and actual risk bounds for minimum variance portfolio construction using our estimator. The maximum exposure bound is of particular importance since it is shared by the theoretical minimum variance portfolio which assumes the integrated covariance matrix is known.

2 Framework and Methodology

2.1 Integrated and realized covariance matrices

Let $X_t = (X_t^{(1)}, \ldots, X_t^{(p)})^{\mathrm{T}}$ be a p-dimensional log-price diffusion process modeled by

$$dX_t = \mu_t dt + \Theta_t dW_t \quad (0 \le t \le 1), \tag{2.1}$$

where μ_t is the drift, Θ_t is a $p \times p$ matrix of instantaneous covolatility processes, and $W_t = (W_t^{(1)}, \ldots, W_t^{(p)})^{\mathrm{T}}$ is a p-dimensional standard Brownian motion. We want to estimate the integrated covariance matrix

$$\Sigma_p = \int_0^1 \Theta_t \Theta_t^{\mathrm{T}} dt. \tag{2.2}$$

This matrix is important in risk management, hedging and pricing of financial derivatives or portfolio allocation, to name but a few areas in finance (Hounyo, 2017). In portfolio allocation, Σ_p replaces the usual population covariance matrix for inter-day data. If Θ_t is constant, then we can take $\Theta_t = \Sigma_p^{1/2}$, and Σ_p is just the usual covariance matrix for asset returns.

In this paper, we consider sparsely sampled return data synchronized by refresh times (Andersen et al., 2001, Barndorff-Nielsen et al., 2011). Suppose we observe X_t at synchronous time points $\tau_{n,\ell}$ ($\ell = 0, \ldots, n$). The realized covariance matrix is then

$$\Sigma_p^{\mathrm{RCV}} = \sum_{\ell=1}^{n} \Delta X_\ell \Delta X_\ell^{\mathrm{T}}, \quad \Delta X_\ell = X_{\tau_{n,\ell}} - X_{\tau_{n,\ell-1}} \quad (\ell = 1, \ldots, n). \tag{2.3}$$

Jacod & Protter (1998) shows that as $n \to \infty$, $\Sigma_p^{\mathrm{RCV}} \to \Sigma_p$ weakly when p is fixed.

2.2 Time-variation adjusted realized covariance matrix

In this section, we present the method of Zheng & Li (2011). Write $dX_t^{(j)} = \mu_t^{(j)} + \sigma_t^{(j)} dZ_t^{(j)}$ ($j = 1, \ldots, p$), where $\mu_t^{(j)}, \sigma_t^{(j)}$ are assumed to be càdlàg over $[0, 1]$, and the $Z_t^{(j)}$'s are one-dimensional standard Brownian motions. Define $\langle X, Y \rangle_t$ to be the quadratic covariation between the processes X and Y.

Assumption 1. *The correlation matrix process of* $Z_t = (Z_t^{(1)}, \ldots, Z_t^{(p)})^{\mathrm{T}}$, $\langle Z^{(j)}, Z^{(k)} \rangle_t / t$ $(j, k = 1, \ldots, p)$, *is constant and non-zero on* $(0, 1]$ *for each* j, k. *Furthermore, the correlation matrix process of* X_t, $\int_0^t \sigma_s^{(j)} \sigma_s^{(k)} d\langle Z^{(j)}, Z^{(k)} \rangle_s \{\int_0^t (\sigma_s^{(j)})^2 ds \int_0^t (\sigma_s^{(k)})^2 ds\}^{-1/2}$ $(j, k = 1, \ldots, p)$, *is constant on* $(0, 1]$ *for each* j, k.

Then, by Proposition 4 of Zheng & Li (2011), there exists a càdlàg process $(\gamma_t)_{t \in [0,1]}$ and a $p \times p$ matrix Λ satisfying $\mathrm{tr}(\Lambda\Lambda^{\mathrm{T}}) = p$ such that we can decompose $\Theta_t = \gamma_t \Lambda$, implying that $\Sigma_p = \big(\int_0^1 \gamma_t^2 dt \big) \Lambda\Lambda^{\mathrm{T}}$. The

time-variation adjusted realized covariance matrix is defined by

$$\check{\Sigma}_p = \frac{\mathrm{tr}(\Sigma_p^{\mathrm{RCV}})}{p}\check{\Phi}, \quad \check{\Phi} = \frac{p}{n}\sum_{\ell=1}^{n}\frac{\Delta X_\ell \Delta X_\ell^{\mathrm{T}}}{\left\|\Delta X_\ell\right\|^2}, \tag{2.4}$$

where $\|\cdot\|$ denotes the L_2 norm of a vector. It is shown in Zheng & Li (2011) that $\mathrm{tr}(\Sigma_p^{\mathrm{RCV}})/p$ is a good estimator for $\int_0^1 \gamma_t^2 dt$, while $\check{\Phi}$ estimates $\Phi = \Lambda\Lambda^{\mathrm{T}}$.

2.3 Nonlinear shrinkage estimator

The estimator $\check{\Phi}$ is a sample covariance matrix of $r_\ell = p^{1/2}\Delta X_\ell/\|\Delta X_\ell\|$ $(\ell = 1, \ldots, n)$, the self-normalized returns. Under the setting $p/n \to c > 0$, the eigenvalues in $\check{\Phi}$ are biased estimators of those in Φ. The way that each r_ℓ $(\ell = 1, \ldots, n)$ is defined means that we cannot apply the nonlinear shrinkage formula in Ledoit & Wolf (2012) directly. Instead, we use the data-splitting idea for nonlinear shrinkage of eigenvalues in Lam (2016).

To this end, we permute the return data M times. At the jth permutation, we split the data $\Delta X^{(j)}$ into two independent parts, say $\Delta X^{(j)} = (\Delta X_1^{(j)}, \Delta X_2^{(j)})$ $(j = 1, \ldots, M)$, with $\Delta X_i^{(j)}$ having size $p \times n_i$ $(i = 1, 2)$ such that $n_1 = m$ and $n_2 = n - m$. Define $\widetilde{\Phi}_i^{(j)} = n_i^{-1}\sum_{\ell \in I_{i,j}} r_\ell r_\ell^{\mathrm{T}}$, where $I_{i,j} = \{\ell : \Delta X_\ell \in \Delta X_i^{(j)}\}$ $(i = 1, 2;\ j = 1, \ldots, M)$. Carrying out an eigen-analysis on $\widetilde{\Phi}_1^{(j)}$, suppose $\widetilde{\Phi}_1^{(j)} = P_1^{(j)} D_1^{(j)} P_1^{(j)\mathrm{T}}$ $(j = 1, \ldots, M)$. Then we introduce our estimator as

$$\widehat{\Sigma}_{m,M} = \frac{\mathrm{tr}(\Sigma_p^{\mathrm{RCV}})}{p}\frac{1}{M}\sum_{j=1}^{M}\widehat{\Phi}^{(j)}, \quad \widehat{\Phi}^{(j)} = P_1^{(j)}\mathrm{diag}(P_1^{(j)\mathrm{T}}\widetilde{\Phi}_2^{(j)}P_1^{(j)})P_1^{(j)\mathrm{T}}, \tag{2.5}$$

where $\mathrm{diag}(\cdot)$ sets all non-diagonal elements of a matrix to 0. The estimator $\widehat{\Phi}^{(j)} (j = 1, \ldots, M)$ belongs to a class of rotation-equivariant estimator $\Phi(D) = P_1^{(j)} D P_1^{(j)\mathrm{T}}$, where D is diagonal. We choose $D = \mathrm{diag}(P_1^{(j)\mathrm{T}}\widetilde{\Phi}_2^{(j)}P_1^{(j)})$ since $\mathrm{diag}(P_1^{(j)\mathrm{T}}\Phi P_1^{(j)})$ solves $\min_D \|P_1^{(j)} D P_1^{(j)\mathrm{T}} - \Phi\|_{\mathrm{F}}$ where $\|A\|_{\mathrm{F}} = \mathrm{tr}^{1/2}(AA^{\mathrm{T}})$, and by Lemma 1, $D^{(j)} = \mathrm{diag}(P_1^{(j)\mathrm{T}}\widetilde{\Phi}_2^{(j)}P_1^{(j)})$ estimates $\mathrm{diag}(P_1^{(j)\mathrm{T}}\Phi P_1^{(j)})$ well. We use the Frobenius norm mainly for the ease of deriving theoretical results. The inverse Stein loss is also considered in Theorem 2.

3 Asymptotic Theory and Practical Implementation

We introduce four more assumptions needed for our results to hold.

Assumption 2. *The drift in (2.1) satisfies $\mu_t = 0$ for $t \in [0, 1]$, and Θ_t is deterministic. All eigenvalues of $\Theta_t\Theta_t^{\mathrm{T}}$ are bounded uniformly from 0 and infinity in $t \in [0, 1]$. Also, M is finite.*

Assumption 3. *The observation times $\tau_{n,\ell}$ are independent of the log-price X_t, and there exists a constant $C > 0$ such that for all positive integer n, $\max_{\ell=1,\ldots,n} n(\tau_{n,\ell} - \tau_{n,\ell-1}) \leq C$.*

Assumption 4. *Let $v_{n,1} \geq \cdots \geq v_{n,p}$ be the p eigenvalues of Φ. Let $H_n(v) = p^{-1} \sum_{i=1}^{p} 1_{\{v_{n,i} \leq v\}}$ be the empirical distribution function of the $v_{n,i}$'s. We assume $H_n(v)$ converges to some non-random limit H at every point of continuity of H.*

Assumption 5. *The support of H defined above is the union of a finite number of compact intervals bounded away from zero and infinity. Also, there exists a compact interval in $(0, +\infty)$ that contains the support of H_n for each n.*

We set $\mu_t = 0$ in Assumption 2 for the ease of proofs and presentation. If μ_t is slowly varying locally, the results to be presented remain valid at the expense of longer proofs. The deterministic nature of Θ_t is essential to the independence of the ΔX_ℓ's. The uniform bounds on the eigenvalues of $\Theta_t \Theta_t^{\mathrm{T}}$ are needed so that individual volatility process for each $X_t^{(i)}$ is bounded uniformly, the integral $\int_0^1 \gamma_t^2 dt > 0$ uniformly, and finally $\|\Sigma_p\| = O(1)$ uniformly. The last two assumptions are essentially Assumptions (A3) and (A4) in Lam (2016) applied on Φ.

Lemma 1. *Let Assumptions 1, 2 and 3 hold for X_t in (2.1). If $p/n \to c > 0$ and $\sum_{n_2 \geq 1} p n_2^{-5} < \infty$, then $\max_{j=1,\ldots,M} \left\| \mathrm{diag}(P_1^{(j)\mathrm{T}} \widetilde{\Phi}_2^{(j)} P_1^{(j)}) \mathrm{diag}^{-1}(P_1^{(j)\mathrm{T}} \Phi P_1^{(j)}) - 1 \right\| \to 0$ almost surely.*

With this result, we can present the following theorem.

Theorem 1. *Let all the assumptions in Lemma 1 hold. Then $\widehat{\Sigma}_{m,M}$ defined in (2.5) is asymptotically almost surely positive definite.*

This is an important result since Σ_p is positive definite, which is not always the case for a realized covariance matrix, especially when $p > n$.

To present the rest of the results, we introduce a benchmark ideal estimator,

$$\Sigma_{\mathrm{ideal}} = \Big(\int_0^1 \gamma_t^2 dt \Big) P \mathrm{diag}(P^{\mathrm{T}} \Phi P) P^{\mathrm{T}}.$$

This is similar to $\widehat{\Sigma}_{m,M}$ defined in (2.5), except that $\mathrm{tr}(\Sigma_p^{\mathrm{RCV}})/p$ is replaced by the population counterpart $\int_0^1 \gamma_t^2 dt$, while $\widehat{\Phi}^{(j)}$ is replaced by $P \mathrm{diag}(P^{\mathrm{T}} \Phi P) P^{\mathrm{T}}$, where P is such that $\check{\Phi} = P \check{D} P^{\mathrm{T}}$, the eigen-decomposition of $\check{\Phi}$ defined in (2.4). Define the efficiency loss of $\widehat{\Sigma}$ as

$$\mathrm{EffLoss}(\Sigma_p, \widehat{\Sigma}) = 1 - \frac{L(\Sigma_p, \Sigma_{\mathrm{Ideal}})}{L(\Sigma_p, \widehat{\Sigma})},$$

where $L(\Sigma_p, \widehat{\Sigma})$ is a loss function. We consider $L(\Sigma_p, \widehat{\Sigma}) = \left\|\widehat{\Sigma} - \Sigma_p\right\|_{\mathrm{F}}^2$, the squared Frobenius loss, and the inverse Stein loss $L(\Sigma_p, \widehat{\Sigma}) = \mathrm{tr}(\Sigma_p \widehat{\Sigma}^{-1}) - \log\det(\Sigma_p \widehat{\Sigma}^{-1}) - p$. If $\widehat{\Sigma}$ incurs a larger loss then Σ_{Ideal}, then $\mathrm{EffLoss}(\Sigma_p, \widehat{\Sigma}) > 0$, and vice versa.

Theorem 2. *Let all the assumptions in Lemma 1 hold, together with Assumptions 4 and 5. Furthermore, if* $n_1/n \to 1$ *and* $n_2 \to \infty$, *then* $\mathrm{EffLoss}(\Sigma_p, \widehat{\Sigma}_p) \leq 0$ *asymptotically almost surely with respect to both the squared Frobenius and the inverse Stein loss functions, provided* $p^{-1}L(\Sigma_p, \Sigma_{Ideal}) \not\to 0$ *almost surely.*

The requirement $p^{-1}L(\Sigma_p, \Sigma_{\mathrm{Ideal}}) \not\to 0$ almost surely eliminates the case $\Sigma_p = (\int_0^1 \gamma_t^2 dt) I_p$, when both the loss functions attain 0 for Σ_{Ideal}. Simulation confirms that $\widehat{\Sigma}_{m,M}$ performs well even in this special case.

To find the best split location m empirically, we minimize

$$g(m) = \Big\| \frac{1}{M} \sum_{j=1}^{M} (\widehat{\Phi}_p^{(j)} - \widetilde{\Phi}_2^{(j)}) \Big\|_{\mathrm{F}}^2.$$

In practice, we use $M = 50$ which provides a good trade-off between computational complexity and estimation accuracy. We search the following split locations for minimizing $g(m)$:

$$m = [2n^{1/2}, 0.2n, 0.4n, 0.6n, 0.8n, n - 2.5n^{1/2}, n - 1.5n^{1/2}].$$

The location $2n^{1/2}$ is suitable for $\Sigma_p = (\int_0^1 \gamma_t^2 dt) I_p$, while $[n - 2.5n^{1/2}]$ and $[n - 1.5n^{1/2}]$ satisfy $\sum_{n_2 \geq 1} p n_2^{-5} < \infty, n_1/n \to 1$ and $n_2 \to \infty$ needed in Theorem 2. We include $0.2n$ to $0.8n$ for boosting finite sample performance.

4 Empirical Results

4.1 Simulations with varying γ_t

In this section, we compare our method to banding (Bickel & Levina, 2008b), the grand average estimator (Abadir et al., 2014), nonlinear shrinkage (Ledoit & Wolf, 2012), principal orthogonal complement thresholding (Fan et al., 2013), the graphical lasso (Friedman et al., 2008), and pure adaptive soft-thresholding. All these methods are applied to $\check{\Phi}$ in (2.4).

Consider two different scenarios for the diffusion process $\{X_t\}$:

Table 1: Mean loss for different methods. Standard errors are subscripted. For the Frobenius loss, all values are multiplied by 10000. The realized covariance matrix is badly conditioned when $n = p = 200$, so the inverse Stein loss does not exist. RCV, realized covariance; Grand Avg, grand average; NONLIN, nonlinear shrinkage; POET, principal orthogonal complement thresholding; GLASSO, graphical lasso; SCAD, adaptive thresholding with the smoothly clipped absolute deviation penalty.

	$p = 100$				$p = 200$			
$n = 200$	Design I Losses		Design II Losses		Design I Losses		Design II Losses	
	Frobenius	Inverse Stein	Frobenius	Inverse Stein	Frobenius	Inverse Stein	Frobenius	Inverse Stein
RCV	94_4	$329.7_{11.7}$	207_9	$271.5_{10.3}$	157_4	–	343_9	–
Proposed	61_3	$18.7_{0.9}$	138_7	$18.7_{0.9}$	83_3	$32.5_{1.2}$	185_6	$32.4_{1.1}$
Banding	73_7	$38.1_{8.7}$	165_{14}	$38.4_{7.9}$	112_{15}	$76.0_{23.1}$	252_{34}	$75.8_{27.4}$
Grand avg	58_3	$10.9_{0.3}$	130_7	$10.9_{0.3}$	76_3	$27.1_{0.6}$	170_6	$27.1_{0.6}$
NONLIN	65_3	$21.5_{1.3}$	147_7	$21.6_{1.3}$	91_3	$134.9_{1032.8}$	204_7	$66.7_{240.4}$
POET	77_3	$11.3_{0.6}$	175_8	$11.4_{0.6}$	112_4	$24.8_{0.9}$	252_8	$24.9_{1.0}$
GLASSO	35_0	$32.1_{0.6}$	79_1	$32.2_{0.6}$	50_0	$64.7_{0.8}$	112_1	$64.7_{0.7}$
SCAD	60_3	$16.2_{0.9}$	135_7	$16.2_{0.9}$	88_3	$54.5_{3.8}$	197_6	$54.9_{3.9}$

Design I: Piecewise constants. We take γ_t to be

$$\gamma_t = \begin{cases} 0.01 \times 7^{1/2} & (0 \le t < 1/4; 3/4 \le t \le 1), \\ 0.01 & (1/4 \le t < 3/4). \end{cases}$$

Design II: Continuous path. We take γ_t to be

$$\gamma_t = \{0.0009 + 0.0008 \cos(2\pi t)\}^{1/2} \quad (0 \le t \le 1).$$

We assume $\Lambda = (0.5^{|i-j|})_{i,j=1,\ldots,p}$, and the observation times are $\tau_{n,\ell} = \ell/n$ $(\ell = 1, \ldots, n)$. We generate $\{X_t\}$ using model (2.1) and get $n = 200$ observations, and consider $p = 100, 200$. For each design and (n, p) combination, we repeat the simulations 500 times , and compare the mean Frobenius and inverse Stein losses for the estimators. We use a five-fold cross-validation to choose the tuning parameter for banding, and $K = 3$ factors for the principal orthogonal complement thresholding with $\theta = 0.5$ as the thresholding parameter, the same as for pure adaptive thresholding. Finally we use $\theta = 0.8$ for the tuning parameter of graphical lasso. These parameters are chosen to allow the methods to have the best possible performances overall. Pre-setting these parameters also speeds up the simulations significantly.

Table 1 presents the simulation results. All methods are better than the realized covariance, as expected. The graphical lasso is the best for minimizing the Frobenius loss, while the grand average estimator at $p = 100$, and the principal orthogonal complement thresholding at $p = 200$, are the best for the inverse Stein loss. Both our method and the grand average estimator outperform nonlinear shrinkage, which is expected since nonlinear shrinkage cannot be readily applied to self-normalized vectors. Although the way that Λ is defined favours banding, it has substantially larger standard deviations in all the settings.

4.2 Portfolio allocation on New York Stock Exchange data

As an application in finance, we construct minimum variance portfolios using the seven different estimators compared in the previous section, except for the graphical lasso because of non-convergence issues. Given an integrated covariance matrix Σ_p, the minimum variance portfolio solves $\min_{w:w^{\mathrm{T}}1_p=1} w^{\mathrm{T}}\Sigma_p w$, where 1_p is a vector of p ones. The solution is

$$w_{\mathrm{opt}} = \frac{\Sigma_p^{-1}1_p}{1_p^{\mathrm{T}}\Sigma_p^{-1}1_p}. \tag{4.6}$$

Before presenting the empirical results, we present a theorem concerning w_{opt} constructed with Σ_p substituted by $\widehat{\Sigma}_{m,M}$. In the sequel, we denote $\|\cdot\|_{\max}$ the maximum absolute value of a vector, and define the condition number of a positive semi-definite matrix A to be $\mathrm{Cond}(A) = \lambda_{\max}(A)/\lambda_{\min}(A)$.

Theorem 3. *Let all the assumptions in Lemma 1 hold. Then, almost surely,*

$$p^{1/2}\|\widehat{w}_{\mathrm{opt}}\|_{\max} \le \mathrm{Cond}(\Phi), \quad p^{1/2}R(\widehat{w}_{\mathrm{opt}}) \le \mathrm{Cond}(\Phi)\lambda_{\max}^{1/2}(\Sigma_p),$$
$$p^{1/2}\|w_{\mathrm{opt}}\|_{\max} \le \mathrm{Cond}(\Phi), \quad p^{1/2}R(w_{\mathrm{opt}}) \le \lambda_{\max}^{1/2}(\Sigma_p),$$

where $\widehat{w}_{\mathrm{opt}}$ is defined in (4.6) with Σ_p substituted by $\widehat{\Sigma}_{p,M}$. The function $R(w) = (w^{\mathrm{T}}\Sigma_p w)^{1/2}$ represents the actual risk when investing using w as the portfolio weights.

This theorem shows that the maximum absolute weight, which we define as the maximum exposure of the portfolio, is decaying at a rate $p^{-1/2}$, the same as that for the actual risk. This maximum exposure bound is important, since the theoretical minimum variance portfolio satisfies the same bound. If $\mathrm{Cond}(\Phi) = 1$, the actual risk for our portfolio can also enjoy the same upper bound as its theoretical counterpart.

We consider $p = 154$ finance stocks with large capitalization from the New York Stock Exchange. There are 82 weeks of data, which starts from June of 2014 to the end of December of 2015. We downloaded all the trades of these stocks from Wharton Research Data Services. The raw data are high-frequency. The stocks have non-synchronous trading times and all the log-prices are contaminated by market microstructure noise (Asparouhova et al., 2013).

We consider trades in 15-minute intervals on every trading day from 9:30 to 16:00, with each log-price being the observed one from a trade right before a 15-minute interval ends. This results in a total of $n = 10267$ synchronized return data points. Overnight returns are not included in all calculations since overnight price jumps are usually influenced by the arrival of news, which is irrelevant to the comparison of portfolios. At the start, we invest 1 unit of capital using (4.6) constructed from different estimators

Table 2: Results for the New York Stock Exchange large capitalization finance stocks analysis. Graphical lasso is omitted because of non-convergence issues. Standard errors are subscripted. All abbreviations are the same as in Table 1.

$p = 154$	Annualized Return(%)	Annualized Std. Dev.(%)	Sharpe Ratio	Maximum Exposure (%)	Max. of Max. Exposure (%)
	Weekly rebalancing with 2-week training windows				
RCV	21.8	12.5	1.7	$25.3_{12.5}$	81.3
Proposed	10.2	9.4	1.1	$7.2_{1.7}$	13.6
Banding	12.5	8.5	1.5	$15.9_{8.7}$	39.2
Grand Avg	10.4	8.9	1.2	$7.4_{2.1}$	14.0
NONLIN	−0.3	8.2	0.0	$5.6_{3.5}$	14.1
POET	−3.9	11.2	−0.3	$19.9_{44.2}$	399.3
SCAD	−15.5	21.2	−0.7	$29.7_{43.6}$	326.3
	Weekly rebalancing with 4-week training windows				
RCV	10.8	11.0	1.0	$20.9_{11.4}$	48.4
Proposed	13.4	9.8	1.4	$8.7_{2.7}$	17.4
Banding	9.3	10.0	0.9	$17.0_{7.5}$	37.6
Grand Avg	11.4	11.1	1.0	$8.0_{1.7}$	13.3
NONLIN	7.6	7.8	1.0	$7.7_{6.0}$	22.8
POET	1.1	11.4	0.1	$20.8_{32.9}$	235.3
SCAD	−4.3	13.7	−0.3	$27.9_{97.1}$	860.6
	Weekly rebalancing with 6-week training windows				
RCV	8.7	8.8	1.0	$19.6_{11.3}$	46.0
Proposed	7.5	10.2	0.7	$10.0_{4.4}$	21.7
Banding	3.7	12.0	0.3	$16.2_{7.5}$	33.6
Grand Avg	2.9	12.2	0.2	$8.7_{2.5}$	14.8
NONLIN	6.8	7.3	0.9	$9.0_{7.4}$	26.1
POET	−9.3	14.4	−0.6	$21.5_{32.5}$	259.6
SCAD	114.9	140.7	0.8	$130.3_{959.1}$	8375.3

of Σ_p. We consider 2-week, 4-week and 6-week training windows and re-evaluate portfolio weights every week. We use the annualized out-of-sample standard deviation $\widehat{\sigma}$, together with the annualized portfolio return $\widehat{\mu}$ and the Sharpe ratio $\widehat{\mu}/\widehat{\sigma}$ to gauge the performance of each method. For ℓ-week training windows and weekly re-evaluation period, $\widehat{\mu}$ and $\widehat{\sigma}$ are defined by

$$\widehat{\mu} = 52 \times \frac{1}{30-\ell} \sum_{i=\ell+1}^{30} w_i^{\mathrm{T}} r_i, \quad \widehat{\sigma} = \Big\{ 52 \times \frac{1}{30-\ell} \sum_{i=\ell+1}^{30} (w_i^{\mathrm{T}} r_i - \widehat{\mu}/52)^2 \Big\}^{1/2} \quad (\ell = 2, 4, 6),$$

where w_i and r_i are portfolio weights and returns respectively for the i-th week. We also report the mean and the maximum of $\big\|\widehat{w}_{\mathrm{opt}}\big\|_{\max}$ over all investment periods for the portfolios constructed under different methods.

Table 2 shows the results. Principal orthogonal complement thresholding and pure adaptive thresholding are unstable, with maximum exposures going over 200% at times, meaning the long or short position on a single stock can be over 200%. This is not practically sound without further information on the stocks. The nonlinear shrinkage method has the smallest $\widehat{\sigma}$ in all settings, followed by our method, banding and grand average. With 6-week training windows, realized covariance has the second smallest $\widehat{\sigma}$, but on average the maximum exposures are much larger than our method and grand average. Our method has small maximum exposures while maintaining Sharpe ratios larger than 0.7 in all settings. It has the

largest Sharpe ratio when we use 4-week training windows.

Supplementary material

Supplementary material includes a set of market trading simulation results and the proof of Lemma 1, Theorem 1, 2 and 3.

References

Abadir, K. M., Distaso, W. & Žikeš, F. (2014). Design-free estimation of variance matrices. *Journal of Econometrics* **181**, 165 – 180.

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* **18**, 351–416.

Andersen, T., Bollerslev, T., Diebold, F. & P., L. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* **96**, 42–55.

Asparouhova, E., Bessembinder, H. & Kalcheva, I. (2013). Noisy prices and inference regarding returns. *The Journal of Finance* **68**, 665–714.

Bai, Z. & Silverstein, J. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. New York: Springer, 2nd ed.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* **162**, 149 – 169.

Bickel, P. J. & Levina, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.* **36**, 2577–2604.

Bickel, P. J. & Levina, E. (2008b). Regularized estimation of large covariance matrices. *Ann. Statist.* **36**, 199–227.

Cai, T. T. & Zhou, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics* **40**, 2389–2420.

Fan, J., Fan, Y. & Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* **147**, 186–197.

Fan, J., Li, Y. & Yu, K. (2012). Vast volatility matrix estimation using high- frequency data for portfolio selection. *Journal of the American Statistical Association* **107**, 412–428.

Fan, J., Liao, Y. & Mincheva, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics* **39**, 3320–3356.

Fan, J., Liao, Y. & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 603–680.

Friedman, J. H., Hastie, T. J. & Tibshirani, R. J. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–441.

Hounyo, U. (2017). Bootstrapping integrated covariance matrix estimators in noisy jump-diffusion models with non-synchronous trading. *Journal of Econometrics* **197**, 130 – 152.

Jacod, J. & Protter, P. (1998). Asymptotic error distributions for the euler method for stochastic differential equations. *Ann. Probab.* **26**, 267–307.

Lam, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *Ann. Statist.* **44**, 928–953.

Lam, C. & Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37**, 4254–4278.

Ledoit, O. & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics* **40**, 1024–1060.

Meinshausen, N. & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* **34**, 1436–1462.

Pourahmadi, M. (2007). Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variancecorrelation parameters. *Biometrika* **94**, 1006–1013.

Rothman, A. J., Bickel, P. J., Levina, E. & Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Statist.* **2**, 494–515.

Zheng, X. & Li, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. *Ann. Statist.* **39**, 3121–3151.