Biometrika (2017), **104**, 2, *pp.* 481–488
Printed in Great Britain
doi: 10.1093/biomet/asx021
Advance Access publication 21 April 2017

Nonlinear shrinkage estimation of large integrated covariance matrices

BY CLIFFORD LAM, PHOENIX FENG AND CHARLIE HU

Department of Statistics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, U.K.

C.Lam2@lse.ac.uk H.Feng2@lse.ac.uk Q.Hu1@lse.ac.uk

SUMMARY

Integrated covariance matrices arise in intraday models of asset returns, which allow volatility to change over the trading day. When the number of assets is large, the natural estimator of such a matrix suffers from bias due to 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. In simulations and a real-data analysis, our estimator performs favourably in comparison with other methods.

Some key words: Extreme eigenvalue; High dimension; Intraday volatility; Maximum exposure bound; Portfolio allocation; Realized covariance.

1. INTRODUCTION

Intraday 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 in estimating this matrix (Aït-Sahalia et al., 2005; Asparouhova et al., 2013). In this paper, we consider the bias that arises when p, the number of assets observed, is large. Specifically, we suppose that p is of 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 can be expressed in terms of the 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, such as banded (Bickel & Levina, 2008b) or sparse covariance matrices (Bickel & Levina, 2008a; Rothman et al., 2008; Lam & Fan, 2009; Cai & Zhou, 2012), sparse precision matrices (Meinshausen & Bühlmann, 2006; Friedman et al., 2008), sparse modified Cholesky factors (Pourahmadi, 2007), 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 assuming a particular structure for the true covariance matrix. However, their approach is not applicable to nonidentically distributed random vectors, such as those arising with intraday return data.

Lam (2016) proved 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 the method of Lam (2016) 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 eigendecompositions of matrices of size $p \times p$, which are not computationally expensive when p is of the order of hundreds, the typical order 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, as it is shared by the theoretical minimum variance portfolio which assumes that 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 modelled by

$$\mathrm{d}X_t = \mu_t \,\mathrm{d}t + \Theta_t \,\mathrm{d}W_t \quad (0 \leqslant t \leqslant 1), \tag{1}$$

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

$$\Sigma_p = \int_0^1 \Theta_t \Theta_t^{\mathrm{T}} \,\mathrm{d}t.$$

This matrix is important in risk management, the hedging and pricing of financial derivatives, and portfolio allocation, to name but a few areas of finance (Hounyo, 2017). In portfolio allocation, Σ_p replaces the usual population covariance matrix for interday 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 that we observe X_t at synchronous time-points $\tau_{n,\ell}$ $(\ell = 0, \ldots, n)$. The realized covariance matrix is then

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

Jacod & Protter (1998) showed 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 $\mathrm{d}X_t^{(j)} = \mu_t^{(j)} + \sigma_t^{(j)} \,\mathrm{d}Z_t^{(j)}$ $(j = 1, \ldots, p)$, where $\mu_t^{(j)}$ and $\sigma_t^{(j)}$ are assumed to be cadlag over $[0, 1]$, and the $Z_t^{(j)}$ 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 nonzero on $(0, 1]$ for each pair of j, k. Furthermore, the correlation matrix process of X_t, $\int_0^t \sigma_s^{(j)} \sigma_s^{(k)} \,\mathrm{d}\langle Z^{(j)}, Z^{(k)}\rangle_s \,\{\int_0^t (\sigma_s^{(j)})^2 \,\mathrm{d}s \int_0^t (\sigma_s^{(k)})^2 \,\mathrm{d}s\}^{-1/2}$ $(j, k = 1, \ldots, p)$, is constant on $(0, 1]$ for each pair of j, k.

Then, by Proposition 4 of Zheng & Li (2011), there exists a cadlag 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 make the decomposition $\Theta_t = \gamma_t \Lambda$, implying that $\Sigma_p = (\int_0^1 \gamma_t^2 \,\mathrm{d}t)\Lambda\Lambda^{\mathrm{T}}$. The time variation-adjusted realized covariance matrix is defined by

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

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 of $\int_0^1 \gamma_t^2\,\mathrm{d}t$, 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 in which each r_ℓ $(\ell = 1,\ldots,n)$ is defined means that we cannot apply the nonlinear shrinkage formula of Ledoit & Wolf (2012) directly. Instead, we use the data-splitting idea for nonlinear shrinkage of eigenvalues from 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 $\tilde{\Phi}_i^{(j)} = n_i^{-1}\sum_{\ell\in I_{i,j}} r_\ell r_\ell^{\mathrm{T}}$, 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 $\tilde{\Phi}_1^{(j)}$, suppose that $\tilde{\Phi}_1^{(j)} = P_1^{(j)} D_1^{(j)} P_1^{(j)\mathrm{T}}$ $(j = 1,\ldots,M)$. Then we define our estimator as

$$\hat{\Sigma}_{m,M} = \frac{\mathrm{tr}(\Sigma_p^{\mathrm{RCV}})}{p}\frac{1}{M}\sum_{j=1}^{M}\hat{\Phi}^{(j)}, \quad \hat{\Phi}^{(j)} = P_1^{(j)}\,\mathrm{diag}(P_1^{(j)\mathrm{T}}\tilde{\Phi}_2^{(j)}P_1^{(j)})P_1^{(j)\mathrm{T}}, \tag{3}$$

where $\mathrm{diag}(\cdot)$ sets all nondiagonal elements of a matrix to zero. The estimator $\hat{\Phi}^{(j)}$ $(j = 1,\ldots,M)$ belongs to a class of rotation-equivariant estimators $\Phi(D) = P_1^{(j)} D P_1^{(j)\mathrm{T}}$, where D is diagonal. We choose $D = \mathrm{diag}(P_1^{(j)\mathrm{T}}\tilde{\Phi}_2^{(j)}P_1^{(j)})$, 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, $D^{(j)} = \mathrm{diag}(P_1^{(j)\mathrm{T}}\tilde{\Phi}_2^{(j)}P_1^{(j)})$ estimates $\mathrm{diag}(P_1^{(j)\mathrm{T}}\Phi P_1^{(j)})$ well. We use the Frobenius norm mainly for ease of deriving theoretical results. In Theorem 2 we also consider the inverse Stein loss.

3. Asymptotic theory and practical implementation

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

Assumption 2. The drift in (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 between zero 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 integers n, $\max_{\ell=1,\ldots,n} n(\tau_{n,\ell} - \tau_{n,\ell-1}) \leqslant C$.

Assumption 4. Let $v_{n,1} \geqslant \cdots \geqslant v_{n,p}$ be the p eigenvalues of Φ. Let $H_n(v) = p^{-1}\sum_{i=1}^{p} 1_{\{v_{n,i}\leqslant v\}}$ be the empirical distribution function of the $v_{n,i}$. We assume that $H_n(v)$ converges to some nonrandom 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 to make the proofs and presentation simpler. If μ_t is slowly varying locally, the results presented here remain valid at the expense of longer proofs. The deterministic nature of Θ_t is essential to the independence of the ΔX_ℓ. The uniform bounds on the eigenvalues of $\Theta_t\Theta_t^{\mathrm{T}}$ are needed so that the individual volatility process for each $X_t^{(i)}$ is bounded uniformly, the integral $\int_0^1 \gamma_t^2\,\mathrm{d}t > 0$ uniformly and, finally, $\|\Sigma_p\| = O(1)$ uniformly. The last two assumptions are essentially assumptions (A3) and (A4) in Lam (2016) applied to Φ.

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

With this result, we have the following theorem.

THEOREM 1. *Let all the assumptions in Lemma 1 hold. Then $\hat{\Sigma}_{m,M}$ defined in (3) is asymptotically almost surely positive definite.*

This is an important result because 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 our results, we introduce a benchmark ideal estimator,

$$\Sigma_{\text{Ideal}} = \left(\int_0^1 \gamma_t^2 \, \mathrm{d}t \right) P \operatorname{diag}(P^{\mathrm{T}} \Phi P) P^{\mathrm{T}}.$$

This is similar to $\hat{\Sigma}_{m,M}$ defined in (3), except that $\operatorname{tr}(\Sigma_p^{\text{RCV}})/p$ is replaced by the population counterpart $\int_0^1 \gamma_t^2 \, \mathrm{d}t$, while $\hat{\Phi}^{(j)}$ is replaced by $P \operatorname{diag}(P^{\mathrm{T}} \Phi P) P^{\mathrm{T}}$, where P is such that $\check{\Phi} = P \check{D} P^{\mathrm{T}}$, the eigendecomposition of $\check{\Phi}$ defined in (2). Define the efficiency loss of $\hat{\Sigma}$ as

$$\operatorname{EffLoss}(\Sigma_p, \hat{\Sigma}) = 1 - \frac{L(\Sigma_p, \Sigma_{\text{Ideal}})}{L(\Sigma_p, \hat{\Sigma})},$$

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

THEOREM 2. *Let all the assumptions in Lemma 1 hold, together with Assumptions 4 and 5. If, moreover, $n_1/n \to 1$ and $n_2 \to \infty$, then $\operatorname{EffLoss}(\Sigma_p, \hat{\Sigma}_p) \leqslant 0$ asymptotically almost surely with respect to both the squared Frobenius and the inverse Stein loss functions, provided that $p^{-1} L(\Sigma_p, \Sigma_{\text{Ideal}}) \not\to 0$ almost surely.*

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

To find the best split location m empirically, we minimize

$$g(m) = \left\| \frac{1}{M} \sum_{j=1}^{M} (\hat{\Phi}_p^{(j)} - \tilde{\Phi}_2^{(j)}) \right\|_{\mathrm{F}}^2.$$

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 $g(m)$:

$$m = \left[2n^{1/2}, 0{\cdot}2n, 0{\cdot}4n, 0{\cdot}6n, 0{\cdot}8n, n - 2{\cdot}5n^{1/2}, n - 1{\cdot}5n^{1/2} \right].$$

The location $2n^{1/2}$ is suitable for $\Sigma_p = (\int_0^1 \gamma_t^2 \, \mathrm{d}t) I_p$, while $[n - 2{\cdot}5n^{1/2}]$ and $[n - 1{\cdot}5n^{1/2}]$ satisfy the conditions $\sum_{n_2 \geqslant 1} p n_2^{-5} < \infty$, $n_1/n \to 1$ and $n_2 \to \infty$ needed in Theorem 2. We include $0{\cdot}2n$ to $0{\cdot}8n$ for boosting finite-sample performance.

4. EMPIRICAL RESULTS

4·1. *Simulations with varying γ_t*

In this section, we compare our method with 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).

Table 1. *Mean losses for different methods, with standard errors in subscript. For the Frobenius loss, all values are multiplied by* 10 000. *The realized covariance matrix is poorly conditioned when $n = p = 200$, so the inverse Stein loss does not exist*

	$p = 100$				$p = 200$			
$n = 200$	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	$329{\cdot}7_{11{\cdot}7}$	207_9	$271{\cdot}5_{10{\cdot}3}$	157_4	–	343_9	–
Proposed	61_3	$18{\cdot}7_{0{\cdot}9}$	138_7	$18{\cdot}7_{0{\cdot}9}$	83_3	$32{\cdot}5_{1{\cdot}2}$	185_6	$32{\cdot}4_{1{\cdot}1}$
Banding	73_7	$38{\cdot}1_{8{\cdot}7}$	165_{14}	$38{\cdot}4_{7{\cdot}9}$	112_{15}	$76{\cdot}0_{23{\cdot}1}$	252_{34}	$75{\cdot}8_{27{\cdot}4}$
Grand Avg	58_3	$10{\cdot}9_{0{\cdot}3}$	130_7	$10{\cdot}9_{0{\cdot}3}$	76_3	$27{\cdot}1_{0{\cdot}6}$	170_6	$27{\cdot}1_{0{\cdot}6}$
NONLIN	65_3	$21{\cdot}5_{1{\cdot}3}$	147_7	$21{\cdot}6_{1{\cdot}3}$	91_3	$134{\cdot}9_{1032{\cdot}8}$	204_7	$66{\cdot}7_{240{\cdot}4}$
POET	77_3	$11{\cdot}3_{0{\cdot}6}$	175_8	$11{\cdot}4_{0{\cdot}6}$	112_4	$24{\cdot}8_{0{\cdot}9}$	252_8	$24{\cdot}9_{1{\cdot}0}$
GLASSO	35_0	$32{\cdot}1_{0{\cdot}6}$	79_1	$32{\cdot}2_{0{\cdot}6}$	50_0	$64{\cdot}7_{0{\cdot}8}$	112_1	$64{\cdot}7_{0{\cdot}7}$
SCAD	60_3	$16{\cdot}2_{0{\cdot}9}$	135_7	$16{\cdot}2_{0{\cdot}9}$	88_3	$54{\cdot}5_{3{\cdot}8}$	197_6	$54{\cdot}9_{3{\cdot}9}$

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.

We consider two scenarios for the diffusion process $\{X_t\}$:

Design I: piecewise constants. We take γ_t to be

$$\gamma_t = \begin{cases} 0{\cdot}01 \times 7^{1/2}, & 0 \leqslant t < 1/4 \text{ or } 3/4 \leqslant t \leqslant 1; \\ 0{\cdot}01, & 1/4 \leqslant t < 3/4. \end{cases}$$

Design II: continuous path. We take γ_t to be

$$\gamma_t = \{0{\cdot}0009 + 0{\cdot}0008 \cos(2\pi t)\}^{1/2} \quad (0 \leqslant t \leqslant 1).$$

We assume that $\Lambda = (0{\cdot}5^{|i-j|})_{i,j=1,\ldots,p}$, and the observation times are $\tau_{n,\ell} = \ell/n$ ($\ell = 1, \ldots, n$). We generate $\{X_t\}$ using model (1), obtaining $n = 200$ observations, and take $p = 100$ or 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 five-fold crossvalidation to choose the tuning parameter for banding, and use $K = 3$ factors for the principal orthogonal complement thresholding with $\theta = 0{\cdot}5$ as the thresholding parameter, the same as for pure adaptive thresholding. Finally, we use $\theta = 0{\cdot}8$ for the tuning parameter of the 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 performed better than the realized covariance, as expected. The graphical lasso was best at minimizing the Frobenius loss, while the grand average estimator with $p = 100$, and the principal orthogonal complement thresholding with $p = 200$, were the best for the inverse Stein loss. Both our method and the grand average estimator outperformed nonlinear shrinkage, which is expected since nonlinear shrinkage cannot readily be applied to self-normalized vectors. Although the way in which Λ is defined favours banding, that method had 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 subsection, except for the graphical lasso because of nonconvergence 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

$$w_{\mathrm{opt}} = \frac{\Sigma_p^{-1} 1_p}{1_p^{\mathrm{T}} \Sigma_p^{-1} 1_p}. \tag{4}$$

Before presenting the empirical results, we state a theorem concerning w_{opt} constructed with $\hat{\Sigma}_{m,M}$ substituted for Σ_p. In what follows, we denote by $\|\cdot\|_{\max}$ the maximum absolute value of a vector, and define the condition number of a positive-semidefinite 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,*

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

where $\hat{w}_{\mathrm{opt}}$ is defined in (4) with Σ_p replaced by $\hat{\Sigma}_{p,M}$. 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 of $p^{-1/2}$, the same as 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, from June 2014 to the end of December 2015. We downloaded all the trades of these stocks from Wharton Research Data Services. The raw data are high-frequency. The stocks have nonsynchronous 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 the end of a 15-minute interval. This results in a total of $n = 10\,267$ synchronized return data points. Overnight returns are not included in the calculations as overnight price jumps are usually influenced by the arrival of news, which is irrelevant to the comparison of portfolios. At the start, we invest one unit of capital using (4) constructed from different estimators of Σ_p. We consider two-week, four-week and six-week training windows and re-evaluate portfolio weights every week. We use the annualized out-of-sample standard deviation $\hat{\sigma}$, together with the annualized portfolio return $\hat{\mu}$ and the Sharpe ratio $\hat{\mu}/\hat{\sigma}$, to gauge the performance of each method. For ℓ-week training windows and a weekly re-evaluation period, $\hat{\mu}$ and $\hat{\sigma}$ are defined by

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

where w_i and r_i are the portfolio weights and returns, respectively, for the ith week. We also report the mean and the maximum of $\|\hat{w}_{\mathrm{opt}}\|_{\max}$ over all investment periods for the portfolios constructed with different methods.

Table 2 shows the results. Principal orthogonal complement thresholding and pure adaptive thresholding are unstable, with maximum exposures going above 200% at times, meaning that 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 $\hat{\sigma}$ in all settings, followed by our method, banding and the grand average estimator. With six-week training windows, realized covariance has the second smallest $\hat{\sigma}$, but on average the maximum exposures are much larger than in our method and the grand average estimator. Our method has small maximum exposures while maintaining Sharpe ratios greater than 0·7 in all settings. It has the largest Sharpe ratio when we use four-week training windows.

Table 2. *Results of the analysis for the New York Stock Exchange large capitalization finance stocks; standard errors are given in subscript. Graphical lasso is omitted because of nonconvergence issues*

$p = 154$	Annualized return (%)	Annualized std dev. (%)	Sharpe ratio	Maximum exposure (%)	Maximum of max. exposure (%)
		Weekly rebalancing with 2-week training windows			
RCV	21·8	12·5	1·7	$25{\cdot}3_{12\cdot5}$	81·3
Proposed	10·2	9·4	1·1	$7{\cdot}2_{1\cdot7}$	13·6
Banding	12·5	8·5	1·5	$15{\cdot}9_{8\cdot7}$	39·2
Grand Avg	10·4	8·9	1·2	$7{\cdot}4_{2\cdot1}$	14·0
NONLIN	−0·3	8·2	0·0	$5{\cdot}6_{3\cdot5}$	14·1
POET	−3·9	11·2	−0·3	$19{\cdot}9_{44\cdot2}$	399·3
SCAD	−15·5	21·2	−0·7	$29{\cdot}7_{43\cdot6}$	326·3
		Weekly rebalancing with 4-week training windows			
RCV	10·8	11·0	1·0	$20{\cdot}9_{11\cdot4}$	48·4
Proposed	13·4	9·8	1·4	$8{\cdot}7_{2\cdot7}$	17·4
Banding	9·3	10·0	0·9	$17{\cdot}0_{7\cdot5}$	37·6
Grand Avg	11·4	11·1	1·0	$8{\cdot}0_{1\cdot7}$	13·3
NONLIN	7·6	7·8	1·0	$7{\cdot}7_{6\cdot0}$	22·8
POET	1·1	11·4	0·1	$20{\cdot}8_{32\cdot9}$	235·3
SCAD	−4·3	13·7	−0·3	$27{\cdot}9_{97\cdot1}$	860·6
		Weekly rebalancing with 6-week training windows			
RCV	8·7	8·8	1·0	$19{\cdot}6_{11\cdot3}$	46·0
Proposed	7·5	10·2	0·7	$10{\cdot}0_{4\cdot4}$	21·7
Banding	3·7	12·0	0·3	$16{\cdot}2_{7\cdot5}$	33·6
Grand Avg	2·9	12·2	0·2	$8{\cdot}7_{2\cdot5}$	14·8
NONLIN	6·8	7·3	0·9	$9{\cdot}0_{7\cdot4}$	26·1
POET	−9·3	14·4	−0·6	$21{\cdot}5_{32\cdot5}$	259·6
SCAD	114·9	140·7	0·8	$130{\cdot}3_{959\cdot1}$	8375·3

All abbreviations are as in Table 1.

Supplementary material

Supplementary material available at *Biometrika* online includes a set of market trading simulation results and the proofs of Lemma 1 and Theorems 1–3.

References

Abadir, K. M., Distaso, W. & Žikeš, F. (2014). Design-free estimation of variance matrices. *J. Economet.* **181**, 165–80.

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

Andersen, T., Bollerslev, T., Diebold, F. & Labys, P. (2001). The distribution of realized exchange rate volatility. *J. Am. Statist. Assoc.* **96**, 42–55.

Asparouhova, E., Bessembinder, H. & Kalcheva, I. (2013). Noisy prices and inference regarding returns. *J. Finan.* **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. *J. Economet.* **162**, 149–69.

Bickel, P. J. & Levina, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.* **36**, 2577–604.

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. *Ann. Statist.* **40**, 2389–420.

FAN, J., FAN, Y. & LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Economet.* **147**, 186–97.

FAN, J., LI, Y. & YU, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *J. Am. Statist. Assoc.* **107**, 412–28.

FAN, J., LIAO, Y. & MINCHEVA, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* **39**, 3320–56.

FAN, J., LIAO, Y. & MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with Discussion). *J. R. Statist. Soc.* B **75**, 603–80.

FRIEDMAN, J. H., HASTIE, T. J. & TIBSHIRANI, R. J. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–41.

HOUNYO, U. (2017). Bootstrapping integrated covariance matrix estimators in noisy jump-diffusion models with non-synchronous trading. *J. Economet.* **197**, 130–52.

JACOD, J. & PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Prob.* **26**, 267–307.

LAM, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *Ann. Statist.* **44**, 928–53.

LAM, C. & FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37**, 4254–78.

LEDOIT, O. & WOLF, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40**, 1024–60.

MEINSHAUSEN, N. & BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34**, 1436–62.

POURAHMADI, M. (2007). Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variance correlation parameters. *Biometrika* **94**, 1006–13.

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**, 3121–51.

[*Received on* 26 *March* 2016. *Editorial decision on* 9 *March* 2017]