Statistics and Its Interface Volume 11 (2018) 141–152

Large portfolio allocation using high-frequency financial data

Jian Zou*, Fangfang Wang, and Yichao Wu

Asset allocation strategy involves dividing an investment portfolio among different assets according to their risk levels. In recent decades, estimating volatilities of asset returns based on high-frequency data has emerged as a topic of interest in financial econometrics. However, most available methods are not directly applicable when the number of assets involved is large, since small component-wise estimation errors could accumulate to large matrix-wise errors. In this paper, we introduce a method to carry out efficient asset allocation using sparsity-inducing regularization on the realized volatility matrix obtained from intraday high-frequency data. We illustrate the new method with the high-frequency price data on stocks traded in New York Stock Exchange over a period of six months in 2013. Simulation studies based on popular volatility models are also presented. The proposed methodology is theoretically justified. Numerical results also show that our approach performs well in portfolio allocation by pooling together the strengths of regularization and estimation from a high-frequency finance perspective.

Keywords and phrases: Portfolio allocation, Risk management, Volatility matrix estimation, High-frequency data, Regularization.

1. INTRODUCTION

Portfolio optimization is the process of determining the optimal mix of assets to hold in the portfolio, which is a very important issue in risk management. Asset allocation strategy involves dividing an investment portfolio among different assets based on the volatilities of the asset returns. The origin of modern portfolio theory goes back to the seminal work of Markowitz [40] which introduced the mean-variance analysis by solving an unconstrained quadratic optimization problem. It was later expanded in the book Markowitz [39]. This approach has had a significant impact on financial economics and is widely used by practitioners. Sharpe [41] introduced the Sharpe ratio to assess the performance of mutual funds, which is a direct measure of reward-to-risk. The theory of portfolio optimization works well if the size of portfolio is small. When the investment portfolio contains a large number of assets, the Markowitz portfolio becomes very sensitive to estimation errors in the expected return and the conditional covariance or volatility matrix of daily returns (see e.g., Chopra and Ziemba [14]).

With widely available high-frequency data in recent years, a broad array of methods have been proposed to model the volatility under this framework. High-Frequency financial data usually refer to intraday observations. The extra amount of information contained in high-frequency data and keen interests in high-frequency finance motivate researchers to develop better estimators and inference procedures for the volatility matrix. The volatility matrix of asset returns plays an important role in portfolio allocation, option pricing, and risk management. However, the main challenge is that the volatility matrix cannot be estimated accurately when the number of assets is large. To fix ideas, suppose we have a portfolio containing 500 stocks. Then we need to estimate $(500^2 + 500)/2 = 125250$ distinct entries simultaneously in the volatility matrix in order to perform asset allocation. Even worse, it has to be done each time portfolio rebalancing occurs. To address this high-dimensional problem, several innovative approaches for volatility matrix estimation are proposed in the literature. Univariate estimation methods include, but are not limited to, realized volatility (Andersen et al. [2]), bi-power realized variation (Barndorff-Nielsen and Shephard [5]), two-time scale realized volatility (Zhang et al. [48]), wavelet realized volatility (Fan and Wang [22]), kernel realized volatility (Barndorff-Nielsen et al. [3]), Fourier realized volatility (Mancino and Sanfelici [38]), pre-averaging estimator (Jacod et al. [30]), and Quasi-maximum likelihood estimator (Xiu [46]). For multiple assets, we face the so-called non-synchronization issue, which refers to a scenario where high-frequency price data are not aligned properly among different assets, due to the fact that they are traded at various mismatched time points. For bivariate cases, Hayashi and Yoshida [28] and Zhang et al. [48] developed two-time scale methods based on the idea of subsampling and averaging to estimate co-integrated volatility of two assets. Barndorff-Nielsen and Shephard [6] discussed estimation of integrated co-volatility for synchronized high-frequency data. Recent developments in multivariate settings can be found in [45, 15, 4, 42, 24, 8, 9], among others.

To solve the portfolio allocation problem in high dimensional settings, Jagannathan and Ma [31] analyzed the impact of weights constraints in large portfolio allocation, and showed that solving the global minimum variance portfolio problem with some constraints on weights is equivalent to

*Corresponding author. E-mail address: jzou@wpi.edu.

using a shrinkage estimate of the covariance matrix. Ledoit and Wolf [33] proposed to shrink the sample covariance matrix toward the identity matrix and illustrate its effectiveness in portfolio selection. Lai et al. [32] employed a Bayesian framework for the mean-variance portfolio optimization. Fan et al. [20] studied portfolio allocation with gross-exposure constraint combining large volatility matrix estimators under different sampling schemes. Hautsch et al. [27] applied a blocked realized kernel with multivariate GARCH models for portfolio optimization. Further developments on this topic include [37, 1, 18].

In recent big data applications, penalized likelihood methods such as LASSO (Tibshirani [43]) have been extensively studied for high dimensional variable selection and regression. A wide array of research is dedicated to the development of new methods for risk management based on regularization (see e.g., Tibshirani [43], Fan and Lv [21]). However, despite its ability to discover the correct sparse representation of the model (Donoho and Huo [17]), the LASSO estimator is in general a biased estimator especially when the true coefficients are relatively large (Zou [49]). Several remedies have been proposed in the literature to discover the sparsity of the true models, while producing consistent estimates for nonzero regression coefficients, such as the smoothly clipped absolute deviation (SCAD) (Fan and Li [19]) and the adaptive LASSO (Zou [49]). Yuan and Lin [47] and Levina et al. [34] proposed shrinkage estimators of the precision matrix through graphical models. Guo et al. [25] extended this methodology to accommodate multiple graphical models. Zou and Wang [50] applied a regularization method in a regression setup. Zou and Zhang [51] explored possible high performance computing solutions to a large dimensional risk problem. In this paper, we adopt the idea of shrinkage estimators and propose a regularized version of the inverse of the average realized volatility matrix estimator of Wang and Zou [45].

The data that motivate this research comprise the tick-by-tick stock prices from the New York Stock Exchange (NYSE). Due to the highly liquidity and vast trading volume, the data set contains high frequency asset price data with some idiosyncratic features, including unequally spaced time intervals, nonsynchronization issue, and price discreteness (see e.g., Wang and Zou [45] for some illustrations of these issues). Figure 1 depicts the sheer size of the one year's trading volume for these stocks. Figure 2 shows the transaction frequency of the 100 stocks in S&P 100 in a day. It demonstrates a U shape pattern indicating heavy trading activities at the opening and closing of the market. The normal trading hours of the NYSE are from 09:30 until 16:00 EST. Therefore, we discard any transactions beyond these hours from our analysis for simplicity.

In this article, we propose a new method to perform efficient portfolio allocation by applying sparsity-inducing regularization on the integrated volatility matrix estimated via intra-day ultra-high-frequency data. The method combines the strengths of high dimensional volatility matrix estimation using high-frequency financial data, and sparsity-inducing regularization techniques that are increasingly popular in the high dimensional variable selection literature. These results may easily be extended to mean-variance and tangency portfolios under the modern portfolio theory. From a financial point of view, the shrinkage estimate of the volatility matrix may be interpreted as some implied volatility matrix with the portfolio manager's view embedded.

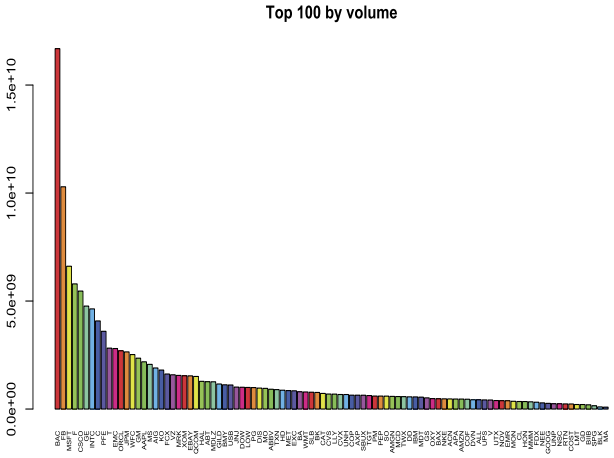

Figure 1. Volume summary of the top 100 traded assets in the New York Stock Exchange in 2013.

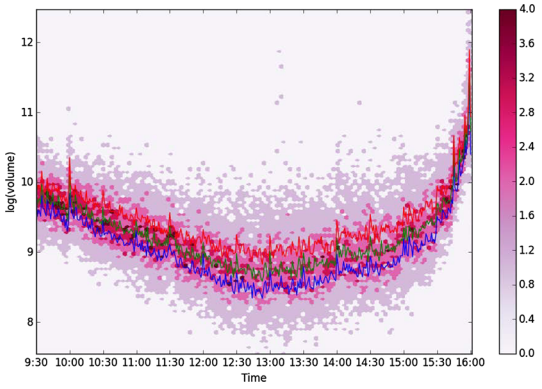

Figure 2. The red line represents the third quantile; the green line represents the second quantile; the blue line represents the first quantile. The color bar represents the frequency of the transaction at any given time. We can clearly see a smiley curve indicating that the trading volume is large at the opening hours and closing hours, while comparatively quiet at noon time.

The rest of this paper is structured as follows. We outline the framework of our method and propose an improved estimator of the inverse volatility matrix in Section 2. The asymptotic properties are discussed in Section 3. Section 4 presents numerical evidence on the performance of our method. Finally, we summarize our findings and offer some concluding remarks in Section 5.

2. METHODS

Consider a portfolio consisting of p assets. The log price $X(t) = (X_1(t), \cdots, X_p(t))^\top$ of the component assets follows an Itô process governed by

$$dX(t) = \boldsymbol{\mu}(t)\, dt + \boldsymbol{\sigma}(t)\, dB(t), \tag{1}$$

where $B(t)$ is a p-dimensional standard Brownian motion, $\boldsymbol{\mu}(t)$ is a drift term taking values in $\mathbb{R}^p$, and $\boldsymbol{\sigma}(t)$ is a spot volatility matrix of size $p \times p$. We assume both $\boldsymbol{\mu}(t)$ and $\boldsymbol{\sigma}(t)$ are continuous in $t \in \mathbb{R}$. The integrated volatility matrix of the assets over the period $[t, t+h]$ is given by

$$\boldsymbol{\Sigma}_{t,t+h} = \int_t^{t+h} \boldsymbol{\sigma}(s)\boldsymbol{\sigma}(s)^\top\, ds. \tag{2}$$

Suppose the allocation vector is $\boldsymbol{w} \in \mathbb{R}^p$ over the same time span. The *ex-post* variation given below

$$R(\boldsymbol{w}, \boldsymbol{\Sigma}) = \boldsymbol{w}^\top \boldsymbol{\Sigma}_{t,t+h} \boldsymbol{w} \tag{3}$$

measures the risk the portfolio is exposed to over $[t, t+h]$.

The classic Markowitz mean-variance portfolio analysis tends to minimize (3) while retaining the expected portfolio return to a target level. However, the estimation error in portfolio mean return could affect the portfolio weights and produce a suboptimal portfolio (see, for instance, Jagannathan and Ma [31]). This prompts us to adopt another popular portfolio strategy: the *global minimum variance portfolio*, which is the minimum risk portfolio with weights that sum to one. The optimal weights are proportional to certain functionals of the inverse integrated volatility matrix, see e.g., Lai et al. [32]. Following Jagannathan and Ma [31] and Fan et al. [23], we consider the following risk optimization, subject to two different constraints:

$$\min \boldsymbol{w}^\top \boldsymbol{\Sigma}_{t,t+h} \boldsymbol{w}, \qquad s.t.\, \|\boldsymbol{w}\|_1 \le c \text{ and } \boldsymbol{w}^\top \mathbf{1} = 1, \tag{4}$$

where $\|\cdot\|_1$ refers to the L_1 norm and c is the gross exposure parameter that characterizes the total exposure allowed in the portfolio. An optimization case with $c = 1$ corresponds to the no-short-sales restriction as in Jagannathan and Ma [31], whereas $c = \infty$ yields the global minimum risk portfolio. For the sake of brevity, we limit our analysis to these two cases. But our methodology can be easily adapted to other choices of the gross exposure level c.

In what follows, we will discuss the estimation of the integrated volatility matrix $\boldsymbol{\Sigma}_{t,t+h}$.

2.1 Average realized volatility matrix for high-frequency financial data

This section provides a review of the averaging realized volatility matrix (ARVM) estimator of Wang and Zou [45] for the integrated volatility matrix. For ease of exposition, we restrict ourselves to $\int_0^1 \boldsymbol{\sigma}(s)\boldsymbol{\sigma}(s)^\top\, ds$, which is simply written as $\boldsymbol{\Sigma}$.

Prices of the assets that make up the portfolio are collected from each transaction and are thus recorded at discrete time points, $\{t_{i,j}, j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, p\}$, where n_i would vary from equity to equity and $0 \le t_{i,j} \le 1$. The observed log price $Y_i(t_{i,j})$ on the i-th asset is a noisy version of its true log price $X_i(t_{i,j})$ and is assumed to obey an additive noise model

$$Y_i(t_{i,j}) = X_i(t_{i,j}) + \varepsilon_i(t_{i,j}), \tag{5}$$

where $\varepsilon_i(t_{i,j})$ represents market microstructure noise at time $t_{i,j}$, and is assumed to be i.i.d. with mean zero and finite fourth moments and to be independent of $X_i(t)$ at all leads and lags.

Let $\boldsymbol{\tau} = \{\tau_r, r = 1, \cdots, m\}$ be the predetermined sampling points on an evenly spaced grid. For the i-th asset, define previous-tick times

$$\tau_{ir} = max\{t_{i\ell} : t_{i\ell} \le \tau_r, \ell = 1, \cdots, n_i\}, \qquad r = 1, \cdots, m. \tag{6}$$

Based on $\boldsymbol{\tau}$ we define realized co-volatility between assets i and j by

$$\widehat{\Sigma}_{ij}(\boldsymbol{\tau}) = \sum_{r=1}^{m} \left[Y_i(\tau_{i,r}) - Y_i(\tau_{i,r-1})\right] \left[Y_j(\tau_{j,r}) - Y_j(\tau_{j,r-1})\right], \tag{7}$$

and the realized volatility matrix by

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\tau}) = \left(\widehat{\Sigma}_{ij}(\boldsymbol{\tau})\right). \tag{8}$$

Let n be the average sample size $n = p^{-1} \sum_{i=1}^p n_i$. We construct $K = [n/m]$ classes of non-overlapping regular grids, $\boldsymbol{\tau}^k = \{n^{-1}(k-1) + \tau_r, r = 1, \ldots, m\}$, $k = 1, \ldots, K$. For instance, consider $\boldsymbol{\tau}$ as a 5-sec samping grid. Then $m = 4680$ if assuming 6.5 trading hours per day and $\tau_r = 4680^{-1} r$. Suppose that the data we collected are 1-sec returns, as a result of which $n_i = 23400$ and $n = 23400$. This yields $K = 5$ non-overlapping grids. They are $\boldsymbol{\tau}^1 = \{\tau_r, r = 1, \ldots, m\}$, $\boldsymbol{\tau}^2 = \{23400^{-1} + \tau_r, r = 1, \ldots, m\}$, ..., $\boldsymbol{\tau}^5 = \{23400^{-1} 4 + \tau_r, r = 1, \ldots, m\}$.

For each grid $\boldsymbol{\tau}^k$, we construct $\widehat{\Sigma}_{ij}(\boldsymbol{\tau}^k)$ as in (7), based on which we define realized co-volatility $\widehat{\Sigma}_{ij} = K^{-1} \sum_{k=1}^K \widehat{\Sigma}_{ij}(\boldsymbol{\tau}^k)$ between assets i and j and realized volatility matrix

$$\widehat{\boldsymbol{\Sigma}} = (\widehat{\Sigma}_{ij}) = K^{-1} \sum_{k=1}^K \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\tau}^k).$$

Set $\boldsymbol{\eta} = diag(\eta_1, \cdots, \eta_p)$, where η_i is the variance of noise $\varepsilon_{i\ell}$. We estimate η_i by

$$\widehat{\eta}_i = \frac{1}{2n_i} \sum_{\ell=1}^{n_i} [Y_i(\tau_{i,\ell}) - Y_i(\tau_{i,\ell-1})]^2,$$

and denote by $\widehat{\boldsymbol{\eta}} = diag(\widehat{\eta}_1, \cdots, \widehat{\eta}_p)$ the estimator of $\boldsymbol{\eta}$. The ARVM estimator of $\boldsymbol{\Sigma}$ is defined by

$$\tilde{\boldsymbol{\Sigma}} = (\tilde{\Sigma}_{ij}) = \widehat{\boldsymbol{\Sigma}} - 2m\widehat{\boldsymbol{\eta}}, \tag{9}$$

that is, we estimate element Σ_{ij} of $\boldsymbol{\Sigma}$ by $\widehat{\Sigma}_{ij}$ for $i \neq j$ and $\widehat{\Sigma}_{ii} - 2m\widehat{\eta}_i$ for $i = j$. The diagonal elements of $\tilde{\boldsymbol{\Sigma}}$ agree with the two-time scale realized volatility (TSRV) of Zhang et al. [48]. Wang and Zou [45] has showed that

$$\|\tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_d = O_P(n^{-1/6}), \qquad d = \{1, 2, \infty\}, \tag{10}$$

after accounting for other random sources in the data, such as price discreteness and nonsynchronization errors. The convergence rate is improved to $O_P(n^{-1/4})$ in Tao et al. [42] that established the optimal minimax risk for estimating large integrated volatility matrix under the subgaussian tail assumption.

2.2 Regularization

When there is a large number of assets involved, the integrated volatility matrix $\boldsymbol{\Sigma}$ is of large size with $2^{-1}p(p+1)$ distinct entries to be estimated. For a moderate to large p, the realized volatility matrix $\tilde{\boldsymbol{\Sigma}}$ may not pose a reasonably well estimator of $\boldsymbol{\Sigma}$. See Bickel and Levina [10, 11], among others. Consequently, the precision matrix acquired by directly inverting $\tilde{\boldsymbol{\Sigma}}$ is usually unreliable and inaccurate. Regularization is needed to reduce the number of effective entries to a reasonable level so that we can concentrate on producing good estimators for these relatively small number of effective entries in the precision matrix.

In Wang and Zou [45], hard-thresholding is adopted to regularize $\tilde{\boldsymbol{\Sigma}}$. Here, we consider penalized estimate of $\boldsymbol{\Sigma}$ and its inverse matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$, the precision matrix. This is inspired by the impact of weight constraints in portfolio risk optimization. In particular, Jagannathan and Ma [31] documented that solving the global minimum variance portfolio problem with constraints on weights is equivalent to using a shrinkage estimate of the integrated volatility matrix. These results apply also to mean-variance and tangency portfolios. From a financial point of view, the shrinkage estimate of integrated volatility matrix can be interpreted as an implied covariance matrix incorporated by the portfolio manager.

The penalized estimation proceeds as follows. The integrated volatility matrix $\boldsymbol{\Sigma}$ is initially estimated by $\tilde{\boldsymbol{\Sigma}} = (\tilde{\sigma}_{ij})$ defined in (9). Under the penalized likelihood framework, we apply the SCAD penalty $p_\lambda(\cdot)$ and solve the following optimization problem

$$\min_{\Omega} - \log|\Omega| + \operatorname{tr}(\tilde{\boldsymbol{\Sigma}}\Omega) + \sum_{i \neq j} p_\lambda(\omega_{ij}), \tag{11}$$

where ω_{ij} represents the $(i,\ j)$-element of Ω. The appeal of (11) is that it yields a sparse estimator of the precision matrix. Note that (11) is not a convex programming due to the non-convexity of p_λ. We next use the local linear approximation algorithm outlined in Zou [49]. It is an iterative algorithm. Denote the solution at the end of the h^{th} iteration by $\widehat{\Omega}^{(h)} = (\widehat{\omega}_{ij}^{(h)})$. The algorithm updates the solution as follows. At iteration $h+1$, we solve a convex optimization problem

$$\min_{\Omega} - \log|\Omega| + \operatorname{tr}(\tilde{\boldsymbol{\Sigma}}\Omega) + \sum_{i \neq j} p'_\lambda(|\omega_{ij}^{(h)}|)|\omega_{ij}| \tag{12}$$

where $p'_\lambda(\cdot)$ is the first derivative, and denote its solution by $\widehat{\Omega}^{(h+1)} = (\widehat{\omega}_{ij}^{(h+1)})$. We repeat this process until convergence is obtained.

Note that (12) can be solved directly by calling the "glasso" package in R with penalty weight $p'_\lambda(|\omega_{ij}^{(t)}|)$ for $i \neq j$ and 0 otherwise. If the initial estimate $\tilde{\boldsymbol{\Sigma}}$ is invertible, we set $\widehat{\Omega}^{(0)} = \tilde{\boldsymbol{\Sigma}}^{-1}$. Otherwise, we may simply make $\widehat{\Omega}^{(0)}$ a zero matrix. For consistency in the selection of nonzero entries in the precision matrix, we follow Lian [35] and employ a Bayesian Information Criterion (BIC) in selecting the tuning parameter λ.

3. ASYMPTOTIC PROPERTIES

In this section, we provide theoretical justification for our proposed shrinkage estimator of the integrated volatility matrix $\boldsymbol{\Sigma} = \int_0^1 \boldsymbol{\sigma}(s)\boldsymbol{\sigma}(s)^\top ds$.

For the i-th asset, we denote by $\{\tau_{i,r,k}, r = 1, 2, \ldots, m\}$ a collection of previous-tick times associated with the grid $\boldsymbol{\tau}^k$ in the spirit of (6), $k = 1, \ldots, K$. Then the union

$$\bigcup_{k=1}^{K} \{\tau_{i,r,k}, r = 1, 2, \ldots, m\} \tag{13}$$

represents all the previous-tick times for the i-th asset. Our technical assumption stated below in (14) ensures that the sets $\{\tau_{i,r,k}, r = 1, \ldots, m\}$, $k = 1, 2, \ldots, K$, are non-overlapping. Sort all the previous-tick times in (13) in ascending order and re-label them as $\{\tau_{i,j}^0, j = 1, \ldots, mK\}$. Because the data under study are tick-by-tick prices from very liquid stocks, we sample the data in calendar time and convert them to 1-sec returns at the very beginning of our analysis. This renders $n_i = 23400$ and $n = 23400$. As a result, $K * m = n$. Further denote by $L(\boldsymbol{\theta})$ the log-likelihood function, and $Q(\boldsymbol{\theta})$ the penalized likelihood function $L(\boldsymbol{\theta}) - n\sum_{j=1}^{d} p_{\lambda_n}(|\theta_j|)$, where $\boldsymbol{\theta} = \operatorname{vech}(\boldsymbol{\Omega})$ the lower triangular part of $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ and $d = p(p+1)/2$. Let

$$\boldsymbol{\theta}_0 = (\theta_{10}, \cdots, \theta_{d0})^\top = (\boldsymbol{\theta}_{1,0}^\top, \boldsymbol{\theta}_{2,0}^\top)^\top$$

be the true value of the parameters. Without loss of generality, assume that $\boldsymbol{\theta}_{2,0} = \mathbf{0}$. We further introduce the following conditions to establish the asymptotic theory.

(A1) Assume that the noise $\varepsilon_i(t_{i,\ell})$ and the diffusion process $\boldsymbol{X}(t)$ in models (1)-(5) are independent; $\varepsilon_i(t_{i,\ell})$, $i = 1, \cdots, p$, $l = 1, \cdots, n_i$ are independent normal with mean zero and $\operatorname{Var}[\varepsilon_i(t_{i,\ell})] = \eta_i \leq \kappa$ for some positive constant κ.

(A2) Assume that each component of the drift $\boldsymbol{\mu}(t)$ has bounded variation, and

$$\max_{1\le i\le p}\max_{0\le t\le 1}E[|\sigma_{ii}(t)|^{\beta}]<\infty,$$
$$\max_{1\le i\le p}\max_{0\le t\le 1}E[|\mu_i(t)|^{2\beta}]<\infty,$$
$$\max_{1\le i\le p}E[|\varepsilon_i(t_{i\ell})|^{2\beta}]<\infty.$$

for some $\beta\ge 2$.

(A3) There exist some constants C_1 and C_2 such that (14)

$$\max_{1\le i\le p}\frac{n_i}{n}<C_1,\quad \max_{1\le i\le p}\max_{1\le \ell\le n_i}|t_{i,\ell}-t_{i,\ell-1}|\le C_2/n.$$

where $n=p^{-1}(n_1+\cdots+n_p)$.

Theorem 3.1. *Assume models (1)-(5) satisfy conditions (A1) - (A3). If* $\max\{|p''_{\lambda_n}(\theta_{j0})|:\theta_{j0}\neq 0\}\to 0$, *then there exists a local maximizer* $\widehat{\boldsymbol{\theta}}$ *of* $Q(\boldsymbol{\theta})$ *such that* $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|=\mathcal{O}_P(e_n+b_n)$ *where* $b_n=\max\{|p'_{\lambda_n}(\theta_{j0})|:\theta_{j0}\neq 0\}$ *and* b_n *converges to 0 as* $\lambda_n\to 0$; $e_n=n^{-1/6}$ *with the microstructure noise present and* $e_n=n^{-1/3}$ *in the absence of the noise (i.e.,* $\eta_i\equiv 0$ *for* $i=1,\ldots,p$*).*

The theorem below shows that this estimator possesses sparsity property, that is, all the zero entries in $\boldsymbol{\theta}$ can be detected simultaneously with probability tending to 1.

Theorem 3.2. *Assume models (1)-(5) satisfy conditions (A1) - (A3). If* $\lim_{n\to\infty}n^{-1}/(e_n\lambda_n)=0$, *and* $\liminf_{n\to\infty}\liminf_{\theta\to 0+}p'_{\lambda_n}(\theta)/\lambda_n>0$, *then our estimator in Theorem 3.1 satisfies*

$$P\left(\widehat{\boldsymbol{\theta}}_2=0\right)\to 1,\quad \text{as } n\to\infty,$$

where $e_n=n^{-1/6}$ *for the case with microstructure noise and* $e_n=n^{-1/3}$ *for the noiseless case.*

4. NUMERICAL STUDIES

4.1 Simulation

In this section, we adopted the simulation models along the line presented in Wang and Zou [45] to illustrate our method. The goal is to compare the performance of our proposed integrated volatility estimator with that of the regular ARVM estimator under different sparsity and noise levels. By sparsity we mean that a large proportion of the elements in the volatility matrix are small. This phenomenon occurs frequently in applications of finance and biological fields. Examples include large dimensional covariance matrices where only a few elements stand out. Sparsity does not necessarily imply that there is little information in the data about the covariance values. On the contrary, if the data are extensive, i.e., n is large, the asymptotic approximation is usually quite accurate. However, in the case when p is also large, it will introduce a lot of difficulties to estimate the covariance matrices precisely. In fact, most of the conventional methods fail in such situations. To illustrate the nature of this effect, we conducted extensive numerical studies of matrix sparsity in this section. We generate the diagonal elements of the spot volatility matrix from four different stochastic volatility models with leverage effect. These four volatility processes are the geometric Ornstein-Uhlenbeck process, the sum of two Cox-Ingersoll-Ross processes (Cox et al. [16] and Barndorff-Nielsen and Shephard [5]), the volatility process in Nelson's GARCH diffusion limit model (Wang [44]), and the two-factor log-linear stochastic volatility process (Huang and Tauchen [29]).

Specifically, the true log price $X(t)$ of p assets is generated from model (1) with zero drift, namely, the diffusion model,

$$dX(t)=\boldsymbol{\sigma}(t)dB(t),\ t\in[0,1],$$

where $B(t)=(B_{1,t},\ldots,B_{p,t})^{\top}$ is a standard p-dimensional Brownian motion. To specify the diagonal elements, $\sigma_{ii}(t)$, of the spot volatility matrix $\boldsymbol{\sigma}(t)$, we choose the following models to form a large volatility matrix with heterogeneous diffuse pattern.

1. Geometric Ornstein-Uhlenbeck (OU) model

$$d\log\sigma_{ii}^2(t)=-0.10\log\sigma_{ii}^2(t)dt+0.25dW_t, \tag{15}$$

where W_t is a standard Brownian motion. The initial value $\sigma_{ii}^2(0)$ is finite and independent of W_t. We take $Corr(B_{j,t},W_{1,t})=-0.62$ for $j=1,\ldots,p$. Note that the stationary distribution of $\log\sigma_{ii}^2(t)$ is $N(0,0.3125)$.

2. Nelson GARCH diffusion model

$$d\sigma_{ii}^2(t)=\left(0.1-\sigma_{ii}^2(t)\right)dt+0.2\sigma_{ii}^2(t)dW_t, \tag{16}$$

where W_t is a standard Brownian motion, and initial value $\sigma_{ii}^2(0)$ is finite and independent of W_t. We take $Corr(B_{j,t},W_t)=-0.5$ for $j=1,\ldots,p$. The stationary distribution is Inverse Gamma(IG) distribution: $\sigma_{ii}^2(t)\sim IG(51,5)$.

3. Superposition of two Cox-Ingersoll-Ross(CIR) processes

$$\sigma_{ii}^2(t)=v_{1,t}+v_{2,t}, \tag{17}$$

where $v_{l,t}$, $l=1,2$, follows a CIR model

$$dv_{l,t}=\lambda_l(\xi_l-v_{l,t})dt+0.743v_{l,t}dW_{l,t} \tag{18}$$

with $\lambda_1=0.043$, $\lambda_2=3.74$, $\xi_1=0.108$, and $\xi_2=0.401$. Because $2\lambda_l\xi_l>\omega_l^2$, the stationary distribution exists and they are

$$v_{1,t}\sim Gamma(0.391,0.552)\text{ and}$$
$$v_{2,t}\sim Gamma(1.453,0.552).$$

We also assume that $Corr(B_{j,t},W_{l,t})=-0.25$ for $j=1,\ldots,p$ and $l=1,2$.

4. Two-factor log-linear stochastic volatility model

$$\sigma_{ii}^2(t) = s\text{-}\exp(0.04v_{1,t} + 1.5v_{2,t} - 1.2), \tag{19}$$

where

$$dv_{1,t} = -0.00137v_{1,t}dt + dW_{1,t}, \tag{20}$$
$$Corr(B_{j,t}, W_{1,t}) = -0.3,$$
$$dv_{2,t} = -1.386v_{2,t}dt + (1 + 0.25v_{2,t})dW_{2,t}, \tag{21}$$
$$Corr(B_{j,t}, W_{2,t}) = -0.3,$$

and

(22)

$$s\text{-}\exp(u) = \begin{cases} \exp(u) & \text{if } u \le \log(1.5); \\ 1.5(1 - \log(1.5) + u^2/\log(1.5))^{1/2} & \text{if } u > \log(1.5). \end{cases}$$

Because $v_{2,t}$ follows an inverse gamma distribution [see below], we truncate the exponential function after $\log(1.5)$ in order to ensure a linear growth after $\log(1.5)$. Note that $v_{1,t}$ from (20) has solution

$$v_{1,t} = v_{1,0}e^{-0.00137t} + \int_0^t e^{-0.00137(t-s)}dW_{1,s},$$

which is coming from an OU process and has stationary distribution

$$v_{1,t} \sim N(0, 364.964).$$

Though $v_{2,t}$ is not GARCH diffusion, $u_{2,t} = v_{2,t} + 4$ is GARCH diffusion by noticing that

$$\begin{aligned} du_{2,t} &= 1.385(4 - u_{2,t})dt + 0.25u_{2,t}dW_{2,t} \\ &= 1.386(4 - u_{2,t})dt + 0.212\sqrt{1.386}u_{2,t}dW_{2,t}. \end{aligned}$$

As a result, $u_{2,t}$ has a stationary distribution $u_{2,t} \sim IG(45.352, 177.408)$. This result would help us to obtain the initial value for $v_{2,t}$, i.e., inverse the number that is drawn from $Gamma(45.352, 177.408)$ and then subtract 4.

The off-diagonal elements of $\boldsymbol{\sigma}(t)$ are tapered off with a tuning (or correlation) process $\kappa(t)$, which is given by

$$\kappa(t) = \frac{\exp\{2u(t) - 1\}}{\exp\{2u(t) + 1\}}, \tag{23}$$

where

$$du(t) = 0.03[0.64 - u(t)]dt + 0.118u(t) + dW^{\kappa}(t), \tag{24}$$

and

$$W^{\kappa}(t) = \sqrt{0.96}W^{c}(t) - 0.2\sum_{j=1}^{p} B_{j,t}/\sqrt{p},$$

and $W^c(t)$ is a standard 1-dimensional Brownian motion independent of $B(t)$. Model (23) is taken from

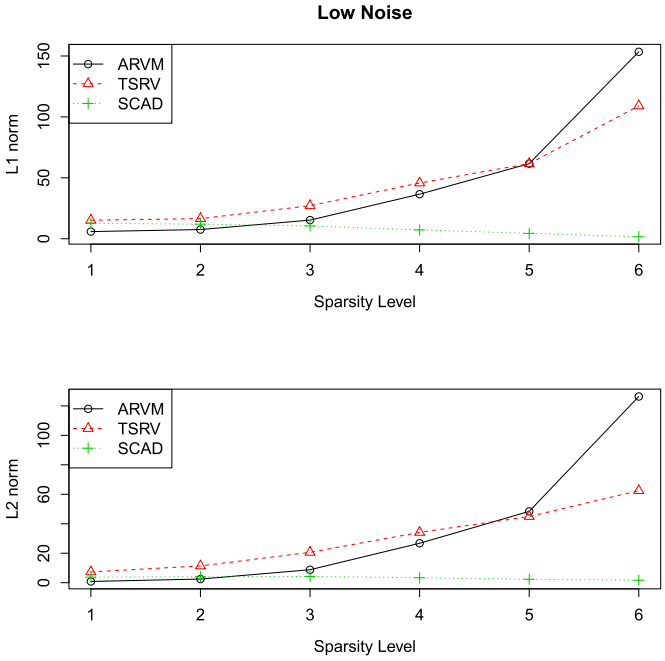

Figure 3. Risk profile with low noise level, smaller risk is better.

Barndorff-Nielsen and Shephard (2004). Because the initial value of the $\kappa(t)$ process determines the decay pattern along the diagonal, we simulate the data generating process $X(t)$ with six different initial values: $\kappa(0) = 0.537, 0.762, 0.905, 0.964, 0.980$, and 0.995. The six cases represent different sparsity levels of the volatility matrices. When $\kappa(0)$ is small, almost all the off-diagonal elements are negligible since the decay of the correlation is very fast. When we increase the initial values and $\kappa(0)$ gets closer to one, the matrices have a much slower decay bands along the diagonal. In particular, $\kappa(0) = 0.537$ is the most concentrated case, and $\kappa(0) = 0.995$ corresponds to the most diffuse case.

In the simulation study we take $p = 512$ assets and $n_j = 200$ intraday observations for $j = 1, \ldots, p$. To mimic the effect of microstructure noise, we add a Gaussian noise with mean zero and variance corresponding to low, medium, and high noise levels to the simulated log prices, following Wang and Zou [45]. We simulate 500 volatility matrices of size 512×512 and calculate the overall risk for the portfolios utilizing ARVM, TSRV and SCAD based volatility estimators, i.e., the matrix L_1 and L_2 error $\|\widehat{\Sigma} - \Sigma\|_d$, $d = 1, 2$. Figures 3-5 reveal that in terms of the portfolio's overall risk based on either the volatility matrix's L_1 norm or L_2 norm, the SCAD penalization based portfolio allocation strategy outperform the ARVM and TSRV based strategies. Although it's worth noting that as the volatility matrix becomes more and more diffuse (or equivalently less and less sparse), the ARVM based strategy performs as

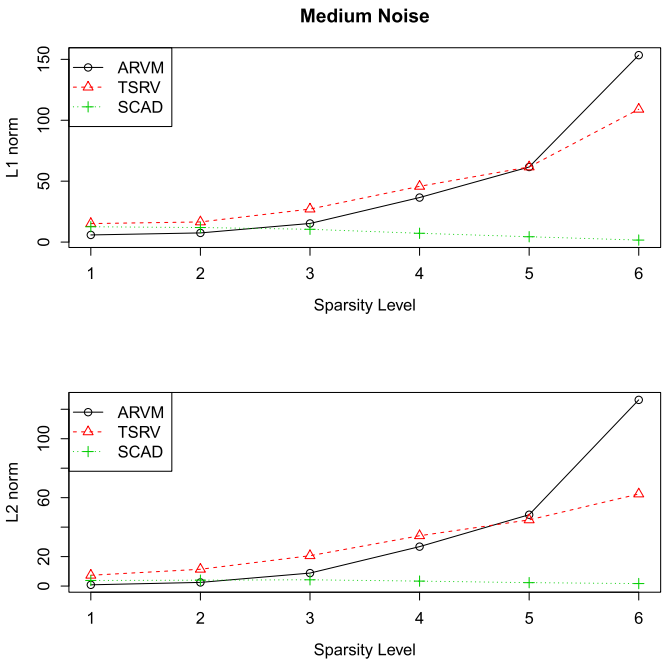

Figure 4. Risk profile with medium noise level.

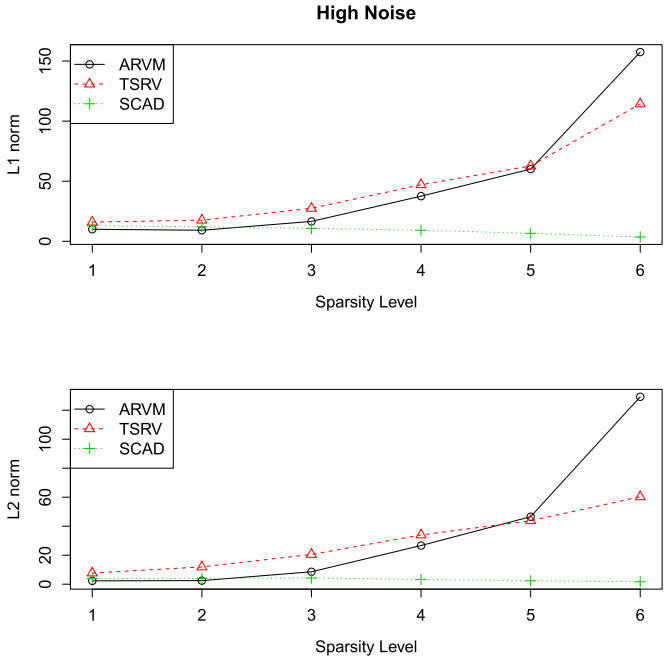

Figure 5. Risk profile with high noise level.

well as our proposed approach. This is because the volatility matrix is no longer sparse and the dependence of the asset components in the portfolio plays an important role in contributing to the overall risk of the total portfolio. Therefore, the advantage of having a sparse estimate of the volatility matrix is less prominent in portfolio performance.

4.2 Daily covariance matrix forecast

The discussion so far explains how to obtain an *ex-post* measure of the integrated volatility matrix $\boldsymbol{\Sigma}_{t,t+h} = \int_t^{t+h} \boldsymbol{\sigma}(s)\boldsymbol{\sigma}(s)^\top ds$ using noisy observations collected between t and $t+h$. To make a sensible economic decision, one is prone to conditional covariance matrix based on which one could calculate *ex-ante* optimal weights. To facilitate exposition, let the unit interval correspond to one day and write $\boldsymbol{\Sigma}_\tau$ for $\boldsymbol{\Sigma}_{\tau-1,\tau}$, $\tau \in \mathbb{Z}$, the daily integrated volatility of day τ. We further define the conditional daily covariance matrix $\boldsymbol{\Sigma}_\tau(1) = E(\boldsymbol{\Sigma}_{\tau+1}|\mathcal{F}_\tau)$ with $\mathcal{F}_t = \sigma(X(s), s \le t)$, which reflects one's belief regarding the future daily variation as of day τ. For a portfolio that is to be held for one day, the asset allocation problem is re-stated using the conditional risk

$$(25) \qquad \min \boldsymbol{w}^\top \boldsymbol{\Sigma}_\tau(1)\boldsymbol{w}, \qquad s.t.\ \|\boldsymbol{w}\|_1 \le c \text{ and } \boldsymbol{w}^\top \mathbf{1} = 1.$$

Multivariate volatility forecasting is always a challenging task in financial econometrics. The conditional covariance $\boldsymbol{\Sigma}_\tau(1)$ is not traceable without knowledge of the volatility matrix's dynamic structure. Denote by $\widehat{\boldsymbol{\Sigma}}_\tau$ the shrinkage estimator of the daily integrated volatility matrix $\boldsymbol{\Sigma}_\tau$. It is natural to consider $E(\widehat{\boldsymbol{\Sigma}}_{\tau+1}|\mathcal{I}_\tau)$ as a proxy for $E(\boldsymbol{\Sigma}_{\tau+1}|\mathcal{F}_\tau)$, where $\mathcal{I}_\tau$ stands for the information up to day τ. Note that $\mathcal{F}_\tau$ is distinct from $\mathcal{I}_\tau$, as the latter pertains to the data that are *observed* at discrete time points. There have been a few attempts in the literature to model daily realized volatility matrices using parsimonious parametric models and to predict future realized measure. Callot et al. [12] proposed to use a vector autoregressive process to model the vast conditional covariance $E(\widehat{\boldsymbol{\Sigma}}_{\tau+1}|\mathcal{I}_\tau)$ where the parameters are estimated via LASSO. Other existing work on multivariate volatility modeling and forecasting includes Chiriac and Voev [13], Bauer and Vorkink [7], Hansen et al. [26], among others.

As vast volatility matrix forecast is not the focus of this paper, we do not resort to sophisticated forecasting models for the realized measure $\widehat{\boldsymbol{\Sigma}}_{\tau+1}$. Let $\widehat{\boldsymbol{\Sigma}}_\tau(1)$ be the estimate of $\boldsymbol{\Sigma}_\tau(1)$. In our empirical exercise, $\widehat{\boldsymbol{\Sigma}}_\tau(1)$ is obtained successively by exponentially weighted moving average. Namely,

$$\widehat{\boldsymbol{\Sigma}}_\tau(1) = (1-\kappa)\widehat{\boldsymbol{\Sigma}}_{\tau-1}(1) + \kappa\widehat{\boldsymbol{\Sigma}}_\tau,$$

and κ can take any value ranging from 0 to 1, which reflects varying degrees of persistence. When κ is 1, this yields the naive forecast. To our pleasant surprise, there are no significant efficiency gains by varying the magnitude of κ in our real data analysis. This is in part due to the nearly stationary behavior of the estimated daily integrated volatility matrix in our setting. Therefore, for simplicity, we choose the naive forecast in the subsequent analysis.

4.3 Real data application

We apply our method to a portfolio consisting of the Standard and Poor (S&P) 100 stocks. The purpose of this

empirical exercise is twofold: to demonstrate the applicability of our approach to a real high-frequency financial data set, as well as to provide insights into regularization in portfolio allocation using high-frequency data.

The S&P 100 index, a sub-set of the S&P 500, measures the performance of large cap companies in the United States. The Index comprises 100 major, blue chip companies across multiple industry groups. Individual stock options are listed for each index constituent. Although changing from time to time, the 100 stocks in the S&P 100 are all major factors in their industries, therefore play an important role in portfolios held by individuals and institutional investors. Due to the changing nature of this index, we use the constituents as of January 1, 2013 to carry out the portfolio allocation. The S&P 100 represents almost one half of the total market capitalization of the U.S., which provides near complete coverage of the U.S. stock market.

We analyze the intraday tick-by-tick trade prices of the S&P 100 stocks from January 1 to June 30 in 2013. The data are extracted from the Trade and Quotes (TAQ) database at Wharton Research Data Services (WRDS) from the Wharton School at the University of Pennsylvania. The data set consists of over 400 million observations across six months and 124 trading days. Since the high-frequency data possess unique features such as nonsynchronous trading and unequally spaced time intervals, we process the data using the same method as in Wang and Zou [45]. To simplify the matter, we only consider the transactions that occurred in the normal trading hours from 9:30 AM to 4:00 PM. Moreover, we ignore the overnight price changes and focus on the impact of volatility matrix on the portfolio performance.

We make asset allocation in accordance with the modern portfolio theory of Markowitz [40, 39]. Two constraints are considered in constructing the portfolio: the no-short-sale constraint (i.e., $c = 1$) and unconstrained weights in the portfolio (i.e., $c = \infty$). The integrated volatility matrix of the 100 stocks is estimated using three different estimators: the ARVM estimator of Wang and Zou [45], the proposed shrinkage estimator that is referred to as SCAD-ARVM, and the multivariate realized kernel (MRK) estimator of [4]. The latter is defined as

$$K_t^b = \sum_{h=-H_t^b}^{H_t^b} k\Big(\frac{h}{H_t^b+1}\Big)\Gamma_t^{h,b}, \tag{26}$$

where $k(x)$ is a Parzen kernel and $\Gamma_t^{h,b}$ is an autocovariance matrix. The smoothing bandwidth H_t^b satisfies $H_t^b \propto n^{(3/5)}$ as suggested in [4]; in practical applications, investors can choose different values according to their own risk profile and the extent of risk tolerance. The flat-top Parzen kernel function is expressed as

$$k(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & 0 \le |x| \le \frac{1}{2} \\ 2(1-|x|)^3 & \frac{1}{2} \le |x| \le 1 \\ 0 & |x| > 1. \end{cases}$$

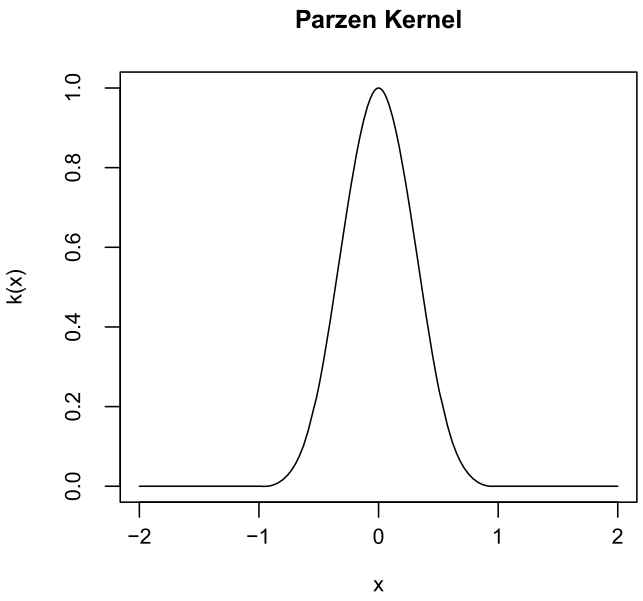

Figure 6. Flat-Top Parzen Kernel function.

To avoid overnight risks due to sudden changes in the price and other potential complications, we restrict our holding period to one trading day in the investment period. The portfolio is then rebalanced daily to maintain the target asset allocation.

At the end of trading day τ, we calculate the optimal weight $\widehat{\boldsymbol{w}}_{\tau+1} = (\widehat{w}_{1,\tau+1}, \ldots, \widehat{w}_{p,\tau+1})^\top$ for the next trading day $\tau + 1$ by solving the optimization problem (25) with $\widehat{\boldsymbol{\Sigma}}_\tau(1)$ in place of $\boldsymbol{\Sigma}_\tau(1)$. Denote by $\boldsymbol{r}_\tau = (r_{1,\tau}, \ldots, r_{100,\tau})^\top$ the daily return of the 100 stocks on day τ. Following Liu [36] and Hautsch et al. [27], we calculate the daily portfolio turnover rate

$$\mathrm{po}_\tau = \sum_{j=1}^{100} \left| \widehat{w}_{j,\tau+1} - \widehat{w}_{j,\tau} \frac{1 + r_{j,\tau}}{1 + \widehat{\boldsymbol{w}}_\tau^\top \boldsymbol{r}_\tau} \right|,$$

which reflects the approximate transaction cost of the associated portfolio allocation strategy due to portfolio rebalancing. The portfolio concentration is measured by the norm of portfolio weight

$$\mathrm{pc}_\tau = \left(\sum_{j=1}^{p} \widehat{w}_{j,\tau}^2 \right)^{1/2},$$

which is minimized for an equally weighted portfolio, i.e., $\widehat{w}_{j,\tau} = m^{-1}$. We also report the size of short positions in the portfolio that is the sum of negative portfolio weights:

$$\mathrm{sp}_\tau = \sum_{j=1}^{p} \widehat{w}_{j,\tau} \mathbf{1}\{\widehat{w}_{j,\tau} < 0\}.$$

Lastly, we introduce the matrix sparsity rate with the aim to investigate whether volatility estimate with a larger

Table 1. Annualized turnover rate (po), portfolio concentration (pc), short position (sp), and sparsity (spar) information of various portfolios constructed using the S&P 100 stocks. m stands for median

	m(po)	m(pc)	m(sp)	m($spar$)
ARVM	0.006	0.183	-0.014	0.996
MRK	0.852	0.318	-0.583	1
SCAD-ARVM	0.421	0.192	-0.176	0.182
SCAD-ARVM (no-short-sale)	0.237	0.182	0	0.033

Table 2. Ex-post performance of various portfolios constructed using the S&P 100 stocks, where σ is the overall risk, sr is the Sharpe ratio, and m stands for median

	m(σ)	SD(σ)	m(sr)	SD(sr)
ARVM	0.056	0.019	0.583	0.525
MRK	0.050	0.016	0.516	0.701
SCAD-ARVM	0.037	0.059	0.636	0.693
SCAD-ARVM (no-short-sale)	0.029	0.092	0.543	0.539

sparsity rate can result in better portfolio performance. The sparsity rate of a n-by-p matrix A is calculated as the number of non-zeros in A divided by its total number of elements:

$$spar = \frac{\sum_{x_i \in A} 1(x_i \neq 0)}{np}.$$

Table 1 reports the median turnover rate, the median portfolio concentration, the median size of short position, and the median sparsity rate. Following the practice in Hautsch et al. [27], we consider median here instead of sample mean. All the statistics are annualized. In contrast to the MRK estimator, the proposed strategy is preferred as it is less costly to implement by comparing their turnover rates. Particularly, the no-short-sale constraint lowers transaction costs. Moreover, the SCAD-ARVM estimator with no-short-sale constraint results in the lowest concentration and thus the most diversified portfolio. The plain ARVM estimator, though yielding a portfolio with the lowest cost, lower concentration, and fewer short positions, is less efficient and less numerically stable than the SCAD-ARVM estimator with no-short-sale constraint due to a high sparsity rate.

To assess the performance of the portfolios, we also compute the global minimum variance and the Sharpe ratio for each trading day. Table 2 reports the median and the standard deviation of the overall risk and the Sharpe ratio for portfolios constructed using ARVM, MRK, and SCAD-ARVM estimators. The proposed shrinkage estimator outperforms the MRK estimator, in that the former results in a lower risk and a higher Sharpe ratio (reward to variability). In addition, the portfolios without short sale constraint pose more risk in the holdings, albeit a higher reward to risk; and the relative risk increase is higher in magnitude than the compensation in the risk-adjusted return. It is interesting as a future direction to further investigate to what extent short sale constraints would attribute to asset allocation since many portfolio managers are prohibited from taking short sale positions.

It is worth noting that the Sharpe ratio should not be used as the only approach to choose investments, rather it is designed for evaluating the performance and risk characteristics of financial assets or funds. In our empirical exercise, we compare portfolios containing the S&P 100 stocks using different asset allocation strategies. Nonetheless, a risk-averse investor may want to invest in a portfolio with a lower Sharpe ratio. Sharpe Ratios should be used to compare investments that fit within the investor's risk tolerance and return profile ([41]).

5. DISCUSSION AND CONCLUSIONS

In this paper, we proposed a new method to carry out efficient asset allocation by applying sparsity-inducing regularization on the integrated volatility matrix estimated via intra-day ultra-high-frequency data. The method combines the strengths of high dimensional volatility matrix estimation using high-frequency financial data, and sparsity-inducing regularization techniques that are increasingly popular in the high dimensional variable selection literature. The regularization-based portfolio selection strategy offers several advantages. First, it can significantly reduce the accumulation of small component-wise estimation errors which can in turn lead to large overall portfolio risk. This helps answer questions about performance and risk as part of a broader investment decision-making process. Second, our framework can easily accommodate different positions such as short sale constraints. Third, by imposing a sparse structure on the integrated volatility matrix, we can improve portfolio allocation efficiency through exploring a smaller asset universe and achieve better numerical stability. We illustrated the proposed method with the high-frequency price data on stocks traded in New York Stock Exchanges over a period of six months in 2013. The results show that our approach performs well in portfolio allocation while pooling together the strengths of regularization and estimation from a high-frequency finance perspective. Our hope is that using such tools can help the decision-maker find proper ways to evaluate the risks of their portfolios at hand.

ACKNOWLEDGMENTS

We thank the editor Professor Heping Zhang, an associate editor, and two reviewers for constructive comments that greatly improved this manuscript. Y. Wu is supported by NSF grant DMS-1055210.

APPENDIX A. PROOFS

In this section, we prove our approach possesses asymptotically consistent and sparsistent properties. Here we follow the price model setup in [45] for large volatility matrix

estimation, and portfolio risk optimization setup as in Jagannathan and Ma [31] and Fan et al. [23].

A.1 Proof of Theorem 1

Proof. Suppose asset price follow the geometric Brownian motion model, the log returns Y_t has iid Gaussian errors. Note that in the penalized estimation setting, the log-likelihood $L(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log|\Omega| - \frac{n}{2}\mathrm{tr}(\Sigma\Omega)$ can be estimated consistently by $L(\widehat{\boldsymbol{\theta}}) = -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log|\widehat{\Omega}| - \frac{n}{2}\mathrm{tr}(\tilde{\Sigma}\widehat{\Omega})$ for any consistent estimator $\tilde{\Sigma}$, i.e., $\|\tilde{\Sigma} - \Sigma\|_d = O_p(e_n)$, $d = \{1, 2, \infty\}$. Therefore, we next show that the penalized estimate $\widehat{\Omega}$ is also a consistent estimate of Ω.

Let $r_n = e_n + b_n$, we want to show that for any given $\varepsilon > 0$, there exists a large constant C such that

$$P\left\{\sup_{\|\boldsymbol{u}\|=C} Q(\boldsymbol{\theta}_0 + r_n\boldsymbol{u}) < Q(\boldsymbol{\theta}_0)\right\} \geq 1 - \varepsilon. \tag{27}$$

This implies that with probability at least $1-\varepsilon$, there exists a local maximum in the ball $\{\boldsymbol{\theta}_0 + r_n\boldsymbol{u} : \|\boldsymbol{u}\| \leq C\}$. Hence, there exists a local maximizer such that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = \mathcal{O}_P(r_n)$.

Since $p_{\lambda_n}(0) = 0$, and $p_{\lambda_n}(|\theta|) \geq 0$ for all θ, we have

(28)

$$\begin{aligned} D_n(\boldsymbol{u}) &\equiv Q(\boldsymbol{\theta}_0 + r_n\boldsymbol{u}) - Q(\boldsymbol{\theta}_0) \\ &\leq L(\boldsymbol{\theta}_0 + r_n\boldsymbol{u}) - L(\boldsymbol{\theta}_0) - n\sum_{j=1}^{s}\left\{p_{\lambda_n}(|\theta_{j0} + r_n u_j|)\right. \\ &\quad \left. - p_{\lambda_n}(|\theta_{j0}|)\right\}, \end{aligned}$$

where s is the number of components in $\boldsymbol{\theta}_{10}$. Let $L'(\boldsymbol{\theta}_0)$ be the gradient vector of L. By Taylor expansion of L, we have

$$\begin{aligned} &D_n(\boldsymbol{u}) \\ &\leq \underbrace{r_n\boldsymbol{u}^\top L'(\boldsymbol{\theta}_0)}_{I_1} \underbrace{-\frac{1}{2}\boldsymbol{u}^\top I(\boldsymbol{\theta}_0)\boldsymbol{u}nr_n^2\{1 + o_P(1)\}}_{I_2} \\ &\underbrace{-n\sum_{j=1}^{s}[r_n\ p'_{\lambda_n}(|\theta_{j0}|\mathrm{sgn}(\theta_{j0})u_j + r_n^2\ p''_{\lambda_n}(|\theta_{j0}|)u_j^2\{1 + o(1)\}]}_{I_3}. \end{aligned}$$

Note that by Theorem 1 of Wang and Zou (2010), we have for all $1 \leq j \leq d$, $\|\widehat{\theta}_j - \theta_0\| \leq Ce_n$. Therefore, I_1 is on the order $O_p(n^{1/6}r_n)$ for the noise case and $O_p(n^{1/3}r_n)$ for the noiseless case. I_2 is on the order $O_p(nr_n^2)$ which is equivalent to $O_p(n^{2/3})$ for the noise case and $O_p(n^{1/3})$ for the noiseless case. By choosing a sufficiently large C, I_1 is uniformly bounded by I_2. For I_3, by Cauchy-Schwartz inequality we have

$$\begin{aligned} I_3 &= -n\sum_{j=1}^{s}[r_n\ p'_{\lambda_n}(|\theta_{j0}|\mathrm{sgn}(\theta_{j0})u_j \\ &\quad + r_n^2\ p''_{\lambda_n}(|\theta_{j0}|)u_j^2\{1 + o(1)\}] \\ &\leq \sqrt{s}nr_na_n\|\boldsymbol{u}\| + nr_n^2\max\{p''_{\lambda_n}(|\theta_{j0}|) : \theta_{j0} \neq 0\}\|\boldsymbol{u}\|^2 \\ &= Cr_n(\sqrt{s}na_n + nr_n\max\{p''_{\lambda_n}(|\theta_{j0}|) : \theta_{j0} \neq 0\}C). \end{aligned}$$

This is also dominated by I_2. Therefore, by choosing a sufficiently large C, (27) holds. Hence, this completes the proof of the theorem. □

A.2 Proof of Theorem 2

Proof. It is sufficient to show that with probability tending to 1 as $n \to \infty$, for any given $\boldsymbol{\theta}_1$ satisfying $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_{10} = O_p(e_n)$ and for some small $\varepsilon_n = Ce_n$ and $j = s+1, \cdots, d$,

$$\frac{\partial Q(\boldsymbol{\theta})}{\partial\theta_j} \leq 0 \quad \text{for } 0 < \theta_j < \varepsilon_n, \tag{29}$$

$$\geq 0 \quad \text{for } -\varepsilon_n < \theta_j < 0. \tag{30}$$

To show (29), note that by Taylor's expansion, we have

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta})}{\partial\theta_j} &= \frac{\partial L(\boldsymbol{\theta})}{\partial\theta_j} - np'_{\lambda_n}(|\theta_j|)\mathrm{sgn}(\theta_j) \\ &= \frac{\partial L(\boldsymbol{\theta}_0)}{\partial\theta_j} + \sum_{l=1}^{d}\frac{\partial^2 L(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\theta_l}(\theta_l - \theta_{l0}) \\ &\quad + \sum_{l=1}^{d}\sum_{k=1}^{d}\frac{\partial^3 L(\boldsymbol{\theta}^*)}{\partial\theta_j\partial\theta_l\theta_k}(\theta_l - \theta_{l0})(\theta_k - \theta_{k0}) \\ &\quad - np'_{\lambda_n}(|\theta_j|)\mathrm{sgn}(\theta_j), \end{aligned}$$

where $\boldsymbol{\theta}^*$ lies between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Note that

$$n^{-1}\frac{\partial L(\boldsymbol{\theta}_0)}{\partial\theta_j} = O_p(n^{-1}/e_n),$$

and

$$n^{-1}\frac{\partial^2 L(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\theta_l} = E[\frac{\partial^2 L(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\theta_l}] + O_p(1).$$

By the assumption that $\boldsymbol{\theta} - \boldsymbol{\theta}_0 = O_p(e_n)$, we have

$$\frac{\partial Q(\boldsymbol{\theta})}{\partial\theta_j} = n\lambda_n\left\{-\lambda_n^{-1}p'_{\lambda_n}(|\theta_j|)\mathrm{sgn}(\theta_j) + O_p(n^{-1}/(e_n\lambda_n))\right\}.$$

Since $\lim_{n\to\infty} n^{-1}/(e_n\lambda_n) \to 0$, and $\liminf_{n\to\infty}\liminf_{\theta\to 0+} p'_{\lambda_n}(\theta)/\lambda_n > 0$, the sign of the derivative is completely determined by that of θ_j. Thus, (29) and (30) follow. This completes the proof of the theorem. □

Received 30 October 2016

REFERENCES

[1] Ait-Sahalia, Y. and Xiu, D. (2016). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*. In Press.

[2] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.*, 96(453):42–55.

[3] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536. MR2468558

[4] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics*, 162(2):149–169. MR2795610

[5] Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(2):253–280. MR1904704

[6] Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925. MR2051439

[7] Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1):93–101. MR2745870

[8] Bibinger, M. (2012). An estimator for the quadratic covariation of asynchronously observed itô processes with noise: Asymptotic distribution theory. *Stochastic Processes and their Applications*, 122(6):2411–2453.

[9] Bibinger, M., Hautsch, N., Malec, P., and Reiss, M. (2014). Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *Annals of Statistics*, 42:1312–1346.

[10] Bickel, P. J. and Levina, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.*, 36(6):2577–2604. MR2485008

[11] Bickel, P. J. and Levina, E. (2008b). Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227. MR2387969

[12] Callot, L. A., Kock, A. B., and Medeiros, M. C. (2016). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*.

[13] Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947. MR2843112

[14] Chopra, V. and Ziemba, W. (1993). The effect of errors in means, variance and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6–11.

[15] Christensen, K., Kinnebrock, S., and Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159(1):116–133.

[16] Cox, J. C., Ingersoll, Jr., J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

[17] Donoho, D. L. and Huo, X. (2002). Beamlets and multiscale image analysis. In *Multiscale and multiresolution methods*, volume 20 of *Lect. Notes Comput. Sci. Eng.*, pages 149–196. Springer, Berlin.

[18] Fan, J., Furger, A., and Xiu, D. (2016). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business & Economic Statistics*, 34(4):489–503. MR3547991

[19] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456):1348–1360.

[20] Fan, J., Li, Y., and Yu, K. (2012a). Vast volatility matrix estimation using high frequency data for portfolio selection. *Journal of the American Statistical Association*, 107(497):412–428.

[21] Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statist. Sinica*, 20(1):101–148.

[22] Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *J. Amer. Statist. Assoc.*, 102(480):1349–1362.

[23] Fan, J., Zhang, J., and Yu, K. (2012b). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606. MR2980070

[24] Griffin, J. E. and Oomen, R. C. (2011). Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *Journal of Econometrics*, 160(1):58–68.

[25] Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika*, 98(1):1–15.

[26] Hansen, P. R., Lunde, A., and Voev, V. (2014). Realized beta GARCH: a multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics*, 29(5):774–799. MR3258063

[27] Hautsch, N., Kyj, L. M., and Malec, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, 30(2):263–290.

[28] Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379.

[29] Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3:456–499.

[30] Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*, 119(7):2249–2276.

[31] Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58:1651–1684.

[32] Lai, T. L., Xing, H., and Chen, Z. (2011). Mean-variance portfolio optimization when means and covariances are unknown. *Ann. Appl. Stat.*, 5(2A):798–823.

[33] Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance*, 10:603–621.

[34] Levina, E., Rothman, A., and Zhu, J. (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *Ann. Appl. Stat.*, 2(1):245–263.

[35] Lian, H. (2011). Shrinkage tuning parameter selection in precision matrices estimation. *J. Statist. Plann. Inference*, 141(8):2839–2848.

[36] Liu, Q. (2009). On portfolio optimization: How and when do we benefit from high-frequency data? *Journal of Applied Econometrics*, 24(4):560–582.

[37] Lunde, A., Shephard, N., and Sheppard, K. (2015). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics*, (just-accepted):1–46.

[38] Mancino, M. E. and Sanfelici, S. (2008). Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Comput. Statist. Data Anal.*, 52(6):2966–2989.

[39] Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, New York.

[40] Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.

[41] Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, 39(1):119–138.

[42] Tao, M., Wang, Y., Yao, Q., and Zou, J. (2011). Large volatility matrix inference via combining low-frequency and high-frequency approaches. *J. Amer. Statist. Assoc.*, 106(495):1025–1040. MR2894761

[43] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288.

[44] Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *Ann. Statist.*, 30(3):754–783. Dedicated to the memory of Lucien Le Cam.
[45] Wang, Y. and Zou, J. (2010). Vast volatility matrix estimation for high-frequency financial data. *Ann. Statist.*, 38(2):943–978.
[46] Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *J. Econometrics*, 159(1):235–250. MR2720855
[47] Yuan, M. and Lin, Y. (2007). Model election and estimation in the gaussian graphical model. *Biometrika*, 94:19–35.
[48] Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472):1394–1411.
[49] Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 101(476):1418–1429.
[50] Zou, J. and Wang, Y. (2013). Statistical methods for large portfolio risk management. *Statistics and Its Interface*, 6:477–485.
[51] Zou, J. and Zhang, H. (2014). High-frequency financial statistics with parallel R and Intel Xeon Phi coprocessor. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 61–69.

Jian Zou
Department of Mathematical Sciences
Worcester Polytechnic Institute
USA
E-mail address: jzou@wpi.edu

Fangfang Wang
Department of Statistics
University of Connecticut
USA
E-mail address: fangfang.wang@uconn.edu

Yichao Wu
Department of Mathematics, Statistics, and Computer Science
University of Illinois at Chicago
Chicago, IL 60607-7045
USA
E-mail address: yichaowu@uic.edu