High-dimensional Minimum Variance Portfolio Estimation Based on High-frequency Data

Tony Cai *
Department of Statistics
University of Pennsylvania

Jianchang Hu †
Department of Statistics
University of Wisconsin-Madison

Yingying Li ‡
Department of ISOM & Department of Finance
HKUST

Xinghua Zheng§
Department of ISOM
HKUST

January 20, 2018

Abstract

This paper studies the estimation of high-dimensional minimum variance portfolio (MVP) based on high frequency returns which can exhibit heteroskedasticity *and* possibly be contaminated by microstructure noise. Under certain sparsity assumptions on the precision matrix, we propose an estimator of MVP and prove that our portfolio asymptotically achieves the minimum variance in a sharp sense. In addition, we introduce consistent estimators of the minimum variance, which provide reference targets. Simulation and empirical studies demonstrate that our proposed portfolio performs favorably.

Key Words: Minimum variance portfolio; High dimension; High frequency; CLIME estimator.

JEL Codes: C13, C55, C58, G11

*Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104. tcai@wharton.upenn.edu

†Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706. huj@stat.wisc.edu

‡Department of ISOM and Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. yyli@ust.hk

§Department of ISOM, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. xhzheng@ust.hk

1 Introduction

1.1 Background

Since the ground-breaking work of Markowitz (1952), the mean-variance portfolio has caught significant attention from both researchers and practitioners. To implement such a strategy in practice, the accuracy in estimating both the expected returns and the covariance structure of returns is vital. It has been well documented that the estimation of the expected returns is more difficult than the estimation of covariances (Merton (1980)), and the impact on portfolio performance caused by the estimation error in the expected returns is larger than that caused by the error in covariance estimation. These difficulties pose serious challenges for the practical implementation of the Markowitz portfolio optimization. The minimum variance portfolio (MVP), on the other hand, avoids the difficulties in estimating the expected returns and is on the efficient frontier with the minimum variance for a given set of assets.

The MVP has received growing attention over the past few years (see, e.g., DeMiguel et al. (2009) and the references therein). In addition to the desirable feature of avoiding mean estimation, it was found to perform well on real data. Empirical studies in Haugen and Baker (1991), Chan et al. (1999), Schwartz (2000), Jagannathan and Ma (2003) and Clarke et al. (2006) have found that the MVP enjoys both lower risks and higher returns compared with some benchmark portfolios. These features make the MVP an attractive investment strategy in practice.

The MVP is more natural in the context of high-frequency data, mostly because the expected returns are negligible over short time horizons[1]. As a prudent common practice, when the time horizon of interest is short, the expected returns are often assumed to be zero (e.g., see Part II of Christoffersen (2012) and the references therein). Fan et al. (2012a) make this assumption when considering the management of portfolios that are rebalanced daily or every other few days. When the expected returns are zero, the mean-variance optimization reduces to the risk minimization problem, in which one seeks the MVP.

There are also benefits of using high-frequency data. On the one hand, large number of observations can potentially help better understand the covariance structure of returns; on the other hand, high-frequency data allows short-horizon rebalancing and hence the portfolios can adjust quickly to time variability of volatilities/co-volatilities. However, high-frequency data do come with significant challenges in analysis. Complications arise due to heteroskedasticity and microstructure noise, among others.

Another interesting problem is the estimation of the global minimum variance. This provides a reference target for the estimated minimum variance portfolios. Moreover, it is useful itself in tracking market risks.

Modern portfolios often involve a large number of assets, and this high-dimensionality poses great challenges. See, for example, Zheng and Li (2011), Fan et al. (2012a), Fan et al. (2012b) and Ao et al. (2014), Xia and Zheng (2017) on issues about and progress made on vast portfolio management.

We consider in this paper the estimation of high-dimensional MVP using high-frequency data. To be more specific, given p assets, which could be p stocks, whose returns $\mathbf{X} =$

[1]For example, as pointed out in Christoffersen (2012), Chapter 1.6, one of the stylized facts of asset returns is that the mean return is usually completely dominated by the volatility over a short period of time such as one day.

$(X_1, \ldots, X_p)^\intercal$ have covariance matrix $\mathbf{\Sigma}$, we aim to find:

$$\arg\min_{\boldsymbol{w}} \boldsymbol{w}^\intercal \mathbf{\Sigma} \boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}^\intercal \mathbf{1} = 1, \tag{1.1}$$

where $\boldsymbol{w} = (w_1, \ldots, w_p)^\intercal$ represents the weights put on different assets, and $\mathbf{1} = (1, \ldots, 1)^\intercal$ is the p-dimensional vector with all entries being 1. The optimal solution is given by

$$\boldsymbol{w}_{\text{opt}} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\intercal \mathbf{\Sigma}^{-1}\mathbf{1}}, \tag{1.2}$$

which yields the minimum risk

$$R_{\min} = \boldsymbol{w}_{\text{opt}}^\intercal \mathbf{\Sigma} \boldsymbol{w}_{\text{opt}} = \frac{1}{\mathbf{1}^\intercal \mathbf{\Sigma}^{-1}\mathbf{1}}. \tag{1.3}$$

More generally, one may be interested in the following optimization problem: for a given vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\intercal$ and c,

$$\arg\min_{\boldsymbol{w}} \widetilde{\boldsymbol{w}}^\intercal \mathbf{\Sigma} \widetilde{\boldsymbol{w}} \quad \text{subject to} \quad \boldsymbol{w}^\intercal \mathbf{1} = c, \text{ where } \widetilde{\boldsymbol{w}} = (\beta_1 w_1, \ldots, \beta_p w_p), \tag{1.4}$$

or its equivalent formulation:

$$\arg\min_{\widetilde{\boldsymbol{w}}} \widetilde{\boldsymbol{w}}^\intercal \mathbf{\Sigma} \widetilde{\boldsymbol{w}} \quad \text{subject to} \quad \widetilde{\boldsymbol{w}}^\intercal \boldsymbol{\beta}^{-1} = c, \text{ where } \boldsymbol{\beta}^{-1} := (1/\beta_1, \ldots, 1/\beta_p)^\intercal. \tag{1.5}$$

Such a setting applies, for example, in leveraged investment. Note that the optimization problem (1.5) can be reduced to (1.1) by noticing that if $\widetilde{\boldsymbol{w}}$ solves (1.5), then $\check{\boldsymbol{w}} :=$

$(\widetilde{w}_1/\beta_1, \ldots, \widetilde{w}_p/\beta_p)^\intercal/c$ solves (1.1) with $\widetilde{\mathbf{\Sigma}} = \text{diag}(\beta_1, \ldots, \beta_p)\mathbf{\Sigma}\,\text{diag}(\beta_1, \ldots, \beta_p)$, and vice versa. For this reason, the two optimization problems (1.1) and (1.4) or (1.5) can be transformed into each other. In the rest of paper we will focus on the optimization problem (1.1).

In practice, because the true covariance matrix is unknown, the sample covariance matrix $\mathbf{S}$ is usually used as a proxy, and the resulting "plug-in" portfolio, $\boldsymbol{w}_p = \mathbf{S}^{-1}\mathbf{1}/\mathbf{1}^\intercal\mathbf{S}^{-1}\mathbf{1}$, has been widely adopted. How well does such a portfolio perform? This question has been considered in Basak et al. (2009). The following simulation result visualizes their first finding (Proposition 1 therein). Figure A.4 shows the risk of the plug-in portfolio based on 100 replications. One can see that the *actual risk* $\widehat{R}_p = \boldsymbol{w}_p^\intercal\mathbf{\Sigma}\boldsymbol{w}_p$ of the plug-in portfolio can be devastatingly higher than the theoretical minimum risk. On the other hand, the *perceived risk* $\widehat{R}_p = \boldsymbol{w}_p^\intercal\mathbf{S}\boldsymbol{w}_p$ can be even smaller than the theoretical minimum risk. Such contradictory phenomena lead to two questions: (1) *Can we consistently estimate the true minimum risk?*; and (2) *More importantly, can we find a portfolio with a risk close to the true minimum risk?*

[*Fig* 1 *here*]

Because of the issue with the plug-in portfolio, alternative methods have been proposed. Jagannathan and Ma (2003) argue that imposing no short-sale constraint helps. More generally, Fan et al. (2012b) study the MVP under the following gross-exposure constraint:

$$\arg\min_{\boldsymbol{w}} \boldsymbol{w}^\intercal\mathbf{\Sigma}\boldsymbol{w} \quad \text{subject to} \quad ||\boldsymbol{w}||_1 \leq \lambda \text{ and } \boldsymbol{w}^\intercal\mathbf{1} = 1, \tag{1.6}$$

where $||\boldsymbol{w}||_1 = \sum_{i=1}^{p} |w_i|$ and λ is a chosen constant. They derive the following bound on the risk of estimated portfolios. If $\widehat{\boldsymbol{\Sigma}}$ is an estimator of $\boldsymbol{\Sigma}$, then the solution to (1.6) with $\boldsymbol{\Sigma}$ replaced by $\widehat{\boldsymbol{\Sigma}}$, denoted by $\widehat{\boldsymbol{w}}_{\text{opt}}$, satisfies that

$$|R_{\min} - R(\widehat{\boldsymbol{w}}_{\text{opt}})| \leq \lambda^2 \cdot ||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_\infty, \tag{1.7}$$

where for any weight vector $\boldsymbol{w}$, $R(\boldsymbol{w}) = \boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}$ stands for the risk measured by the variance of the portfolio return, and for any matrix $A = (a_{ij})$, $||A||_\infty := \max_{ij} |a_{ij}|$. In particular, $||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_\infty$ is the maximum element-wise estimation error in using $\widehat{\boldsymbol{\Sigma}}$ to estimate $\boldsymbol{\Sigma}$. Fan et al. (2012a) consider the high-frequency setting, where they use the two-scale realized covariance matrix (Zhang et al. (2005)) to estimate the so-called integrated covariance matrix (see Section 2.1 below for related background), and establish concentration inequalities for element-wise estimation errors. These concentration inequalities imply that even if the number of assets p grows faster than the number of observations n, one still has that $||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_\infty \to 0$ as $n \to \infty$; see equation (20) therein for the precise statement. In particular, bound (1.7) guarantees that under the gross-exposure constraint, the difference between the risk associated with $\widehat{\boldsymbol{w}}_{\text{opt}}$ and the minimum risk is asymptotically negligible.

The difference between the risk of an estimated portfolio and the minimum risk going to zero, however, may not be as attractive as it sounds. In fact, under rather general assumptions (which do not exclude factor models), the minimum risk $R_{\min} = 1/\mathbf{1}^\intercal \boldsymbol{\Sigma}^{-1} \mathbf{1}$ goes to zero as the number of assets $p \to \infty$; see Ding et al. (2017) for a thorough discussion. If indeed the minimum risk goes to 0 as $p \to \infty$, then the difference $|R_{\min} - R(\widehat{\boldsymbol{w}}_{\text{opt}})| \to 0$ is not enough to guarantee (near) optimality. Based on the above consideration, we turn to find an asset allocation $\widehat{\boldsymbol{w}}$ which satisfies a stronger sense of consistency in that the ratio between the risk of the estimated portfolio and the minimum risk goes to one, i.e.,

$$\frac{R(\widehat{\boldsymbol{w}})}{R_{\min}} \to 1 \quad \text{as} \quad p \to \infty. \tag{1.8}$$

1.2 Main contributions of this paper

Our contributions mainly lie in the following aspects.

We propose a new perspective to study the minimum variance portfolio, namely, to work with ratio consistency (1.8). To achieve the stronger convergence (1.8), we introduce a new approach to estimate the MVP (1.1). It is shown that, under some sparsity assumptions on the inverse of the covariance matrix (also known as the precision matrix), our estimated portfolio enjoys the desired convergence (1.8).

We also introduce a consistent estimator of the minimum risk that does not depend on the sparsity assumption and further establish the related CLT.

Moreover, to utilize high-frequency data, we develop methods to remove impacts due to heteroschadasticity and market microstructure noise, and establish the statistical properties of the estimated minimum variance portfolio.

1.3 Organization of the paper

The paper is organized as follows. In Section 2, we present our estimator of the MVP and show that its risk converges to the minimum risk in the sense of (1.8). We establish the desired convergence in the high-frequency setting, in which case returns may exhibit heteroskedasticity and possibly be contaminated by microstructure noise. A consistent estimator of the minimum risk is proposed in Section 3, for which we also establish the CLT. Section 4 presents simulation results to illustrate the performance of both portfolio and minimum risk estimations. Empirical analysis results based on NYSE stocks are reported in Section 5. We conclude our paper with a brief discussion in Section 6. All the proofs are given in the Appendix.

2 Estimating the MVP

2.1 High-frequency Data

We assume that the latent log-price process $(\mathbf{X}_t)$ follows a diffusion model:

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Theta}_t d\mathbf{W}_t, \quad \text{for} \quad t \in [0,1], \tag{2.1}$$

where $(\boldsymbol{\mu}_t) = (\mu_t^1, \ldots, \mu_t^p)^\intercal$ is the drift process, $(\boldsymbol{\Theta}_t) = (\theta_t^{ij})_{1\le i,j\le p}$ is a $p \times p$ matrix-valued process called the spot co-volatility process, and $(\mathbf{W}_t)$ is a p-dimensional Brownian motion. Both $(\boldsymbol{\mu}_t)$ and $(\boldsymbol{\Theta}_t)$ are stochastic, càdlàg, and may depend on $(\mathbf{W}_t)$, all defined on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0})$. The time interval $[0,1]$ stands for the period of interest. For example, if one plans to rebalance portfolio daily, then the time interval $[0,1]$ represents one day.

Let

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Theta}_t \boldsymbol{\Theta}_t^\intercal := (\sigma_t^{ij})$$

be the spot covariance matrix process. The *ex-post* integrated covariance (ICV) matrix is

$$\mathbf{\Sigma}_{ICV} = (\sigma^{ij}) := \int_0^1 \mathbf{\Sigma}_t dt.$$

Denote its inverse by $\mathbf{\Omega} := \mathbf{\Sigma}_{ICV}^{-1}$. The ex-post minimum risk, $R_{\min}$, is obtained by replacing the $\mathbf{\Sigma}$ in (1.3) with $\mathbf{\Sigma}_{ICV}$.

Let us emphasize that in general, $\mathbf{\Sigma}_{ICV}$ is a random variable which is only measurable to $\mathcal{F}_1$, and so is $R_{\min}$. It is therefore in principle impossible to construct a portfolio that is measurable to $\mathcal{F}_0$ to achieve the minimum risk $R_{\min}$. Practical implementation of the minimum variance portfolio relies on making forecasts about $\mathbf{\Sigma}_{ICV}$ based on historical data. The simplest approach is to assume that $\mathbf{\Sigma}_{ICV,1} \approx \mathbf{\Sigma}_{ICV,2}$, where $\mathbf{\Sigma}_{ICV,i}$ stands for the ICV matrix in period $[i-1, i]$. Under such an assumption, if we can construct a portfolio $\boldsymbol{w}$ based on the observations during $[0, 1]$ (and hence only measurable to $\mathcal{F}_1$) that can approximately minimize the ex-post risk $\boldsymbol{w}^\intercal \mathbf{\Sigma}_{ICV,1} \boldsymbol{w}$, then if we hold the portfolio during the next period $[1, 2]$, the *ex-ante* risk $\boldsymbol{w}^\intercal \mathbf{\Sigma}_{ICV,2} \boldsymbol{w}$ is still approximately minimized. In this article we shall adopt such a strategy.

2.2 High-frequency case with no microstructure noise

We first consider the case when there is no microstructure noise, in other words, one observes the true log-prices (X_t^i).

Our approach to estimate the minimum variance portfolio relies on the constrained l_1-minimization for inverse matrix estimation (CLIME) proposed in Cai et al. (2011). The CLIME estimator $\widehat{\mathbf{\Omega}}$ of $\mathbf{\Omega} := \mathbf{\Sigma}^{-1}$ is defined as the solution to the following optimization problem:

$$\arg\min_{\boldsymbol{\Omega}'} ||\boldsymbol{\Omega}'||_1 \text{ subject to } ||\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Omega}' - \mathbf{I}||_\infty \leq \lambda, \tag{2.2}$$

where $\widehat{\boldsymbol{\Sigma}}$ is an estimator of $\boldsymbol{\Sigma}$, $\mathbf{I}$ is the $p \times p$ identity matrix, and for any matrix $A = (a_{ij})$, $||A||_1 := \sum_{i,j} |a_{ij}|$ and $||A||_\infty := \max_{ij} |a_{ij}|$, and λ is a tuning parameter.

With the CLIME estimator $\widehat{\boldsymbol{\Omega}}$, the resulting estimated MVP is given by

$$\widehat{\boldsymbol{w}} = \frac{\widehat{\boldsymbol{\Omega}}\mathbf{1}}{\mathbf{1}^\intercal\widehat{\boldsymbol{\Omega}}\mathbf{1}}. \tag{2.3}$$

The risk associated with the portfolio $\widehat{\boldsymbol{w}}$ is

$$R_n = \widehat{\boldsymbol{w}}^\intercal\boldsymbol{\Sigma}\widehat{\boldsymbol{w}} = \frac{(\widehat{\boldsymbol{\Omega}}\mathbf{1})^\intercal\boldsymbol{\Sigma}(\widehat{\boldsymbol{\Omega}}\mathbf{1})}{(\mathbf{1}^\intercal\widehat{\boldsymbol{\Omega}}\mathbf{1})^2}. \tag{2.4}$$

If the true log-prices are observed, one of the most commonly used estimators for $\boldsymbol{\Sigma}_{ICV}$ is the realized covariance (RCV) matrix. For each stock i, suppose the observations at stage n are $X^i_{t^{i,n}_\ell}$, where $0 = t^{i,n}_0 < t^{i,n}_1 < \cdots < t^{i,n}_{N_i} = 1$ are the observation times. The n characterizes the observation frequency, and $N_i \to \infty$ as $n \to \infty$. The synchronous observation case corresponds to

$$t^{i,n}_\ell \equiv t^n_\ell \quad \text{for all } i = 1, \ldots, p, \tag{2.5}$$

which, in the simplest equidistant setting, reduces to

$$t^{i,n}_\ell = t^n_\ell = \ell/n, \quad \ell = 0, 1, \ldots, n. \tag{2.6}$$

In the synchronous observation case (2.5), let

$$\Delta \mathbf{X}_\ell := \mathbf{X}_{t_\ell^n} - \mathbf{X}_{t_{\ell-1}^n},\ \ \ell = 1, \ldots, n$$

be the log-return vectors over the time interval $[t_{\ell-1}^n, t_\ell^n]$, then the RCV matrix is defined as

$$\widehat{\mathbf{\Sigma}}_{RCV} = \sum_{\ell=1}^{n} \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^\intercal. \tag{2.7}$$

For any $0 \le q < 1$, $s_0 = s_0(p) < \infty$ and $M = M(p) < \infty$, define a uniformity class of matrices as

$$\mathcal{U}(q, s_0, M) = \Big\{ \mathbf{\Omega} = (\mathbf{\Omega}_{ij})_{p\times p} : \mathbf{\Omega} \text{ positive definite, } ||\mathbf{\Omega}||_{L_1} \le M, \max_{1\le i\le p} \sum_{j=1}^{p} |\mathbf{\Omega}_{ij}|^q \le s_0 \Big\}, \tag{2.8}$$

where $||\mathbf{\Omega}||_{L_1} := \max_{1\le j\le p} \sum_{i=1}^{p} |\mathbf{\Omega}_{ij}|$.

The CLIME estimator is shown in Cai and Liu (2011) to be consistent when the observations are i.i.d. sub-Gaussian, and the underlying $\mathbf{\Omega}$ belongs to $\mathcal{U}(q, s_0(p), M(p))$ with $q, s_0(p)$ and $M(p)$ satisfying $M^{2-2q} s_0 \left(\log p/n\right)^{(1-q)/2} \to 0$; see Theorem 1(a) therein.

We remark that the sparsity assumption $\mathbf{\Omega} = \mathbf{\Sigma}^{-1} \in \mathcal{U}(q, s_0, M)$ appears to be reasonable in financial applications. For example, if the returns are assumed to follow a (conditional) multivariate normal distribution with covariance matrix $\mathbf{\Sigma}$, then the (i, j)th element in $\mathbf{\Omega}$ being 0 is equivalent to that the returns of the i-th and j-th stocks are conditionally independent given the other stock returns. For stocks in different sectors, many pairs might be conditionally independent or only weakly dependent. Practically, as can be seen from the performance of our estimated portfolios in the empirical studies, this assumption appears to be fine.

We derive our theoretical results under the following

Assumption:

A There exists $\delta \in (0,1)$ such that for all p, $0 < \delta \le \lambda_{\min} \le \lambda_{\max} < 1/\delta$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of $\mathbf{\Sigma}_{ICV}$, respectively.

B The drift processes are such that $|\mu_t^i| \le C_\mu$ for some constant $C_\mu < \infty$ for all $i = 1, \ldots, p$ and $t \in [0,1]$ almost surely.

C There exists constant C_σ such that $|\sigma_t^{ij}| \le C_\sigma < \infty$ for all $i = 1, \ldots, p$ and $t \in [0,1]$ almost surely.

D The observation times t_ℓ^n under the synchronous setting (2.5) satisfy that

$$\sup_n \max_{1 \le \ell \le n} n|t_\ell^n - t_{\ell-1}^n| \le C_\Delta < \infty. \tag{2.9}$$

E $p, n \to \infty$ and $\log p/n \to 0$.

Theorem 1. *Suppose that* $(\mathbf{X}_t)$ *satisfies* (2.1) *and the underlying precision matrix* $\mathbf{\Omega} = \mathbf{\Sigma}_{ICV}^{-1} \in \mathcal{U}(q, s_0, M)$. *Under Assumptions A-E, with* $\lambda = \eta M \sqrt{\log p/n}$ *for some* $\eta > 0$, *there exist constants* $C_0, C_1 > 0$ *such that*

$$P\left(\frac{R_n}{R_{\min}} - 1 = O\left(M^{2-2q} s_0 \left((\log p/n)^{(1-q)/2}\right)\right)\right) \ge 1 - \frac{C_0}{p^{C_1\eta^2 - 2}},$$

where $R_n := \widehat{\boldsymbol{w}}^\intercal \mathbf{\Sigma}_{ICV} \widehat{\boldsymbol{w}}$ *is the risk associated with the portfolio* $\widehat{\boldsymbol{w}} = \frac{\widehat{\mathbf{\Omega}}^{RCV}\boldsymbol{1}}{\boldsymbol{1}^\intercal \widehat{\mathbf{\Omega}}^{RCV}\boldsymbol{1}}$, $\widehat{\mathbf{\Omega}}^{RCV}$ *is the*

CLIME estimator obtained by replacing $\widehat{\boldsymbol{\Sigma}}$ *with* $\widehat{\boldsymbol{\Sigma}}_{RCV}$ *in* (2.2)*, and* $R_{\min} = 1/(\boldsymbol{1}^{\intercal}\boldsymbol{\Omega}\boldsymbol{1})$.

Remark 1. *Theorem 1 guarantees that as long as* $M^{2-2q}s_0\left((\log p/n)^{(1-q)/2}\right) \to 0$*, the risk of our estimated MVP is consistent in the sense of* (1.8)*. The estimated portfolio is therefore applicable to the ultra-high-dimensional setting where the number of assets can be much larger than the number of observations.*

Remark 2. *The case where observations are i.i.d. is actually a special case of the high frequency setting that we adopt here. In fact, to generate i.i.d. returns under our setting, one just needs to take constant drift and volatility processes. Our setting is therefore more general than the i.i.d. observation setting, and all our results readily apply to that case.*

Remark 3. *Although Theorem 1 is stated for synchronous observation time case, by using data synchronization methods and following the techniques in Fan et al. (2012a), Theorem 1 can be easily generalized to the asynchronous setting.*

2.3 High-frequency case with microstructure noise

In general, the observed prices are believed to be contaminated by microstructure noise. In other words, instead of observing the true log-prices $X^i_{t^{i,n}_\ell}$, for each stock i, the observations at stage n are

$$Y^i_{t^{i,n}_\ell} = X^i_{t^{i,n}_\ell} + \varepsilon^i_\ell, \tag{2.10}$$

where ε^i_ℓ's represent microstructure noise. In this case, if one simply plugs $Y^i_{t^{i,n}_\ell}$ into the formula of RCV in (2.7), the resulting estimator is not consistent even when the dimension p is fixed. Consistent estimators in the univariate case include the two-scales realized volatility (TSRV, Zhang et al. (2005)), multi-scale realized volatility (MSRV, Zhang (2006)), pre-averaging estimator (PAV, Jacod et al. (2009) and Podolskij and Vetter (2009)), realized kernels (RK, Barndorff-Nielsen et al. (2008)), and quasi-maximum likelihood estimator (QMLE, Xiu (2010)). All these estimators, however, are not consistent in the high-dimensional setting.

In this article we choose to work with the PAV estimator. To reduce non-essential technical complications, we work with the equidistant time setting (2.6) (such as data sampled every minute). Asynchrony can be dealt with by using data synchronization techniques such as the previous tick method (Zhang (2010)) and the refresh time scheme (Barndorff-Nielsen et al. (2011)). Compared with microstructure noise, asynchrony is less an issue, see, for example, Section 2.4 in Xia and Zheng (2017).

To implement the PAV estimator, we fix a $\theta > 0$, and let $k_n = [\theta n^{1/2}]$ be the window length over which the averaging takes place. Define

$$\overline{\mathbf{Y}}_k^n = \frac{\sum_{i=k_n/2}^{k_n-1} \mathbf{Y}_{t_{k+i}} - \sum_{i=0}^{k_n/2-1} \mathbf{Y}_{t_{k+i}}}{k_n}.$$

The PAV (Jacod et al. (2009)) with weight function $g(x) = x \wedge (1-x)$ for $x \in (0,1)$ is defined as

$$\widehat{\mathbf{\Sigma}}_{PAV} = \frac{12}{\theta\sqrt{n}} \sum_{k=0}^{n-k_n+1} \overline{\mathbf{Y}}_k^n \cdot (\overline{\mathbf{Y}}_k^n)^\intercal - \frac{6}{\theta^2 n} \operatorname{diag}\left(\sum_{k=1}^{n} (\Delta Y_{t_k}^i)^2\right)_{i=1,\ldots,p}. \tag{2.11}$$

Now we state additional assumptions that we need in order to establish the statistical results concerning the estimation of MVP based on the PAV estimator.

Assumption:

F The microstructure noise ε_ℓ^i is strictly stationary and independent with mean zero and also independent of $(\mathbf{X}_t)$.

G For any $\theta \in \mathbb{R}$,

$$E(\exp(\theta\varepsilon_\ell^i)) \leq \exp(C\theta^2),$$

where C is a fixed constant. Suppose also that there exists $C_\varepsilon > 0$ such that $\mathrm{Var}(\varepsilon_\ell^i) = \sigma^{ii} \leq C_\varepsilon$ for all i.

H $p, n \to \infty$ and $\log p/\sqrt{n} \to 0$.

Theorem 2. *Suppose that* $(\mathbf{X}_t)$ *satisfies* (2.1) *and underlying precision matrix* $\mathbf{\Omega} = \mathbf{\Sigma}_{ICV}^{-1} \in \mathcal{U}(q, s_0, M)$. *Under Assumptions A-D and F-H, with* $\lambda = \eta M\sqrt{\log p}/n^{1/4}$ *for some* $\eta > 0$, *there exist constant* $C_2, C_3 > 0$ *such that*

$$P\left(\frac{R_n}{R_{\min}} - 1 = O\left(M^{2-2q}s_0\left((\log p)^{(1-q)/2}/n^{(1-q)/4}\right)\right)\right) \geq 1 - \frac{C_2}{p^{(C_3\eta^2-2)}},$$

where R_n *is the risk associated with the portfolio* $\widehat{\boldsymbol{w}} = \frac{\widehat{\mathbf{\Omega}}^{PAV}\boldsymbol{1}}{\boldsymbol{1}^\intercal\widehat{\mathbf{\Omega}}^{PAV}\boldsymbol{1}}$, *and* $\widehat{\mathbf{\Omega}}^{PAV}$ *is the CLIME estimator obtained by replacing* $\widehat{\mathbf{\Sigma}}$ *with* $\widehat{\mathbf{\Sigma}}_{PAV}$ *in* (2.2).

Remark 4. *Theorem 2 states that in the noisy case, if* $M^{2-2q}s_0\left((\log p)^{(1-q)/2}/n^{(1-q)/4}\right) \to 0$, *then the risk of our estimated MVP is consistent in the sense of* (1.8).

Remark 5. *The reduction in the rate from* $(\sqrt{\log p}/\sqrt{n})^{1-q}$ *in Theorem 1 to* $(\sqrt{\log p}/n^{1/4})^{1-q}$ *in the current theorem is an inevitable consequence due to noise. In fact, the optimal rate in estimating the integrated volatility in the noisy case is* $O(n^{1/4})$ *(Gloter and Jacod (2001)), as is compared with the rate of* $O(n^{1/2})$ *in the noiseless case.*

3 Estimating the Minimum Risk

So far we have seen that under certain sparsity assumption on the precision matrix, we can construct a portfolio with a risk close to the true minimum risk in the sense of (1.8). Now we turn to consistent estimation of the minimum risk.

3.1 Assuming Sparsity: Using CLIME based Estimator

The CLIME estimator we considered above also leads to an estimator of the true global minimum risk $R_{\min}$. More specifically, define

$$\widehat{R}_{\mathrm{CLIME}} = \frac{1}{\mathbf{1}^\intercal \widehat{\mathbf{\Omega}} \mathbf{1}}. \tag{3.1}$$

Here, when working under high-frequency setting without noise, $\widehat{\mathbf{\Omega}}$ is the CLIME estimator obtained by replacing $\widehat{\mathbf{\Sigma}}$ with $\widehat{\mathbf{\Sigma}}_{RCV}$ in (2.2), denoted as $\widehat{\mathbf{\Omega}}^{RCV}$. The corresponding minimum risk estimator is defined as

$$\widehat{R}_{\mathrm{CLIME}}^{RCV} = \frac{1}{\mathbf{1}^\intercal \widehat{\mathbf{\Omega}}^{RCV} \mathbf{1}}. \tag{3.2}$$

Similarly, if microstructure noise is present, then $\widehat{\mathbf{\Sigma}}_{PAV}$ is adopted in forming the $\widehat{\mathbf{\Omega}}$, which results in the estimator $\widehat{\mathbf{\Omega}}^{PAV}$. The corresponding minimum risk estimator is then defined as

$$\widehat{R}_{\mathrm{CLIME}}^{PAV} = \frac{1}{\mathbf{1}^\intercal \widehat{\mathbf{\Omega}}^{PAV} \mathbf{1}}. \tag{3.3}$$

We have the following results for these CLIME-based estimators.

Theorem 3. *(i) Under the assumptions in Theorem 1, the minimum risk estimator* $\widehat{R}_{\mathrm{CLIME}}^{RCV}$

defined in (3.2) satisfies that

$$P\left(\frac{\widehat{R}^{RCV}_{\text{CLIME}}}{R_{\min}}-1=O\left(M^{2-2q}s_0\left((\log p/n)^{(1-q)/2}\right)\right)\right)\geq 1-\frac{C_0}{p^{C_1\eta^2-2}}.$$

(ii) Under the assumptions in Theorem 2, the minimum risk estimator $\widehat{R}^{PAV}_{\text{CLIME}}$ *defined in (3.3) satisfies that*

$$P\left(\frac{\widehat{R}^{PAV}_{\text{CLIME}}}{R_{\min}}-1=O\left(M^{2-2q}s_0\left((\log p)^{(1-q)/2}/n^{(1-q)/4}\right)\right)\right)\geq 1-\frac{C_2}{p^{(C_3\eta^2-2)}}.$$

We remark that as long as $M(p)^{2-2q}s_0(p)\left((\log p/n)^{(1-q)/2}\right)\to 0$ in the noise-less case or $M(p)^{2-2q}s_0(p)\left((\log p)^{(1-q)/2}/n^{(1-q)/4}\right)\to 0$ in the noisy case, we would have a consistent estimation of the minimum risk. The estimators are hence applicable to the ultra-high-dimensional setting where the number of assets can be much larger than the number of observations.

3.2 Without the Sparsity Assumption: Low-frequency I.I.D. Returns

CLIME estimator works under mild sparsity assumptions on the precision matrix. Below we propose another estimator of the minimum risk, which does not rely on such sparsity assumption, although on the other hand, it assumes that the observations are i.i.d. normally distributed. This estimator is hence more suitable for the low frequency setting and can be used to estimate the minimum risk during a long time period.

More specifically, suppose that we observe n i.i.d. returns $\mathbf{X}_1,\ldots,\mathbf{X}_n$ (possibly in low frequency). Let $\mathbf{S}$ be the sample covariance matrix, and let $\boldsymbol{w}_p=\mathbf{S}^{-1}\mathbf{1}/\mathbf{1}^\intercal\mathbf{S}^{-1}\mathbf{1}$ be the "plug-in" portfolio weights obtained by replacing $\mathbf{\Sigma}$ in (1.2) with $\mathbf{S}$. The corresponding *perceived* risk is $\widehat{R}_p = \boldsymbol{w}_p^\intercal \mathbf{S} \boldsymbol{w}_p$. We have the following result about the relationship between $\widehat{R}_p$ and the minimum risk $R_{\min}$, based on which a consistent estimator of the minimum risk is constructed.

Theorem 4. *Suppose that the returns $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim_{i.i.d.} N(\boldsymbol{\mu}, \mathbf{\Sigma})$. Suppose further that both n and $p \to \infty$ in such a way that $\rho_n := p/n \to \rho \in (0,1)$. Then*

$$\left| \frac{\widehat{R}_p}{R_{\min}} - (1-\rho_n) \right| \to_p 0. \tag{3.4}$$

Therefore, if we define

$$\widehat{R}_{\min} = 1/(1-\rho_n) \cdot \widehat{R}_p, \tag{3.5}$$

then

$$\frac{\widehat{R}_{\min}}{R_{\min}} \to_p 1. \tag{3.6}$$

Furthermore, we have

$$\sqrt{n-p} \left(\frac{\widehat{R}_{\min}}{R_{\min}} - 1 \right) \Rightarrow N(0,2). \tag{3.7}$$

The convergence (3.7) actually shows "blessing" of dimensionality: the higher the dimension, the more accurate the estimation.

The convergence (3.4) explains why in Figure A.4 the perceived risk is systematically lower than the minimum risk. We also remark that the "plug-in" portfolio is not optimal. In Basak et al. (2009), it is shown that the risk of the plug-in portfolio is *on average* a higher-than-one multiple of the minimum risk; see Propositions 1 & 2 therein. Under the condition that both p and $n \to \infty$ and $p/n \to \rho \in (0,1)$, their result can be strengthened to be that the risk of the plug-in portfolio is with probability approaching one, a larger-than-one multiple of the minimum risk. In fact, using the relationship (5) in Basak et al. (2009), it is easy to show that

$$\frac{R(\boldsymbol{w}_p)}{R_{\min}} \to_p \frac{1}{1-\rho}, \tag{3.8}$$

where $R(\boldsymbol{w}_p) = \boldsymbol{w}_p^\intercal \boldsymbol{\Sigma} \boldsymbol{w}_p$ is the risk of the plug-in portfolio.

4 Simulation Studies

4.1 Case I: when there is no microstructure noise

We first consider a setting in which the log-price process $(\mathbf{X}_t)$ can be observed at high frequeny. We assume that (X_t^i) follows

$$dX_t^i = \left(\mu_t^i - \frac{\sum_{j=1}^p a_{ij}^2 (\sigma_t^j)^2}{2}\right) dt + \sum_{j=1}^p a_{ij}\sigma_t^j dW_t^j, \quad \text{for } t \in [0,1] \text{ and } i = 1, \ldots, p \tag{4.1}$$

where (μ_t^i) and $A = (a_{ij})$ are to be specified, σ_t^j's are such that their logarithms $\varrho_t^j := \log \sigma_t^j$ follow independent Ornstein-Uhlenbeck processes:

$$d\varrho_t^j = \alpha^j (\beta_0^j - \varrho_t^j) dt + \beta_1^j \, dU_t^j, \quad j = 1, \ldots, p, \tag{4.2}$$

where $(\boldsymbol{U}_t) = (U_t^1, \ldots, U_t^1)$ is a p-dimensional standard Brownian motion independent of $(\boldsymbol{W}_t = (W_t^j))$.

Now we specify $A = (a_{ij})$. It is taken to be such that its inverse is a tri-diagonal Toeplitz matrix:

$$A = B^{-1}, \quad \text{where} \quad B = \begin{pmatrix} \alpha & \beta & 0 & 0 & \cdots & 0 \\ \beta & \alpha & \beta & 0 & \cdots & 0 \\ 0 & \beta & \alpha & \beta & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \beta & \alpha \end{pmatrix}. \tag{4.3}$$

Elementary calculations show that if $\alpha^2 \neq 4\beta^2$, then

$$a_{ij} = (-1)^{i+j}\beta^{j-i}\frac{(c_1^i - c_2^i)(c_1^{p+1-j} - c_2^{p+1-j})}{(c_1 - c_2)(c_1^{p+1} - c_2^{p+1})}, \quad 1 \leq i \leq j \leq p,$$

where $c_1 = \frac{\alpha+\sqrt{\alpha^2-4\beta^2}}{2}$ and $c_2 = \frac{\alpha-\sqrt{\alpha^2-4\beta^2}}{2}$; and if $\alpha^2 = 4\beta^2$, then

$$a_{ij} = (-1)^{i+j}\beta^{j-i}\left(\frac{\alpha}{2}\right)^{i-j-1}\frac{i(p-j+1)}{p+1}, \quad 1 \leq i \leq j \leq p.$$

In the simulations below, we fix $\alpha = 3$ and $\beta = -1$.

Under such a setting, the spot co-volatility matrix for $(\mathbf{X}_t)$ is

$$\mathbf{\Sigma}_t = A\widetilde{\mathbf{\Sigma}}_t A,$$

where $\widetilde{\mathbf{\Sigma}_t} = \operatorname{diag}\big((\sigma_t^i)^2\big)$, and the ICV matrix is

$$\text{ICV} = A\int_0^1 \widetilde{\mathbf{\Sigma}}_t\; dt\; A = A\left(\operatorname{diag}\left(\int_0^1 (\sigma_t^i)^2 dt\right)\right)A.$$

In the simulations below, we fix the number of stocks to be $p = 70$. For each day, we simulate 79 observed prices at times ℓ/n, $\ell = 0, 1, \ldots, n$, for $n = 78$. This corresponds to sub-sampling every 5 minutes during the regular trading hours from 9:30 a.m. to 4 p.m.. The number of days is set to be $T = 252$, the average number of trading days in a year. The parameters in (4.2) are taken to be $(\alpha^j, \beta_0^j, \beta_1^j) \sim (\text{Unif}[1.5, 3], \text{Unif}[\log(0.025), \log(0.12)], 10^{-2})$. With these configurations, the simulated average daily volatility will be around 1.5% to 5.5%, similar to the levels observed in real data.

We consider using estimated weights to form a portfolio for the next day (except for the infeasible optimal portfolio, which uses the ICV matrix of the next day). We compare the performance of different portfolios in terms of the annualized standard deviations of their daily returns. The initial portfolio value is set to be \$100 for all portfolios.

We consider the setting with non-zero $(\boldsymbol{\mu}_t)$, in which case the price process is not a martingale. The $(\mu_t^i) \equiv \mu^i$ are sampled independently from Uniform$[-5.67e-6,\ 5.67e-6]$. The resulting annualized returns fall between -10% and 10%. Figure 2 below shows the time series plots of four portfolio values:

(i) Oracle portfolio. This is the optimal portfolio obtained by plugging the ICV matrix into the weight formula (1.2). This portfolio is *infeasible* because ICV is only measurable to the filtration at the end of the day while the portfolio is to be held from the beginning of each day. It is represented by the black curve;

(ii) CLIME-based portfolio. This is our proposed portfolio computed using (2.3) with $\widehat{\boldsymbol{\Omega}}$ obtained by replacing $\widehat{\boldsymbol{\Sigma}}$ with $\widehat{\boldsymbol{\Sigma}}_{RCV}$ in (2.2). The portfolio is represented by the red curve;

(iii) Equally-weighted portfolio. We include this widely used benchmark portfolio for comparison and it is represented by the purple curve;

(iv) Plug-in portfolio. This is obtained by replacing $\mathbf{\Sigma}$ in (1.2) with the RCV matrix (2.7). It is represented by the orange curve.

[*Fig* 2 *here*]

Table 1 reports the annualized standard deviations of daily returns of the compared portfolios.

[*Table* 1 *here*]

From Figure 2 and Table 1 we see that our CLIME-based portfolio closely tracks the infeasible optimal portfolio and has a risk close to the minimum risk. In contrast, the equally-weighted and especially the plug-in portfolios carry much higher risks.

We further apply $\widehat{R}_{\text{CLIME}}^{RCV}$ to estimate the daily minimum risks. As we discussed in Section 2.1, the daily minimum risk is a random variable that changes from one day to the other. Hence, we obtain one estimated minimum risk for each day, and totally 252 estimates. The mean and standard deviation of the estimated minimum risks are 0.471% and 0.027%, respectively. By the law of total variance, under the stationarity assumption, the unconditional daily minimum risk is estimated to be $\sqrt{0.00472^2 + 0.00027^2} \approx 0.472\%$, which corresponds to an annualized risk of 7.50%. This is close to the risk of the oracle portfolio, confirming the validity of our minimum risk estimator.

Alternatively, we can use daily returns to estimate minimum risk based on Theorem 4.

Specifically, we treat daily returns as i.i.d. observations and use the estimator in (3.5). The estimated daily minimum risk is 0.522%, which corresponds to an annualized risk of 8.28%, again similar to the risk of the oracle portfolio.

4.2 Case II: when there is microstructure noise

Next, we consider a setting where the prices are contaminated with noise. More specifically, we adopt the same setting for the true log-price process $(\mathbf{X}_t)$ with the parameters generated in the same way as in Case I. In addition, we simulate noise process independently from $N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon \sim \text{Unif}[9e-4, 3.6e-3]$.

We still consider $p = 70$ stocks. For each day, we simulate 391 observed prices at equally spaced times. This corresponds to sub-sampling every minute during the regular trading hours. The number of days is again set to be 252. Figure 3 below shows the time series plots of the following portfolio values:

(i) Oracle portfolio. Again we include this *infeasible* optimal portfolio in our comparison. It is represented by the black curve;

(ii) 5-Min CLIME-based portfolio. This is obtained based on Theorem 1 using 5-min returns and is represented by red curve;

(iii) 1-Min-PAV-CLIME-based portfolio. This is based on Theorem 2 with pre-averaging estimator using 1-min returns and is represented by blue curve;

(iv) Equally-weighted portfolio. We again include this benchmark portfolio and is represented by the orange curve;

(v) Plug-in portfolio. The portfolio based on RCV is also constructed using 5-min returns.

It is represented by the purple curve.

[*Fig* 3 *here*]

Table 2 reports the corresponding annualized standard deviations.

[*Table* 2 *here*]

Figure 3 and Table 2 show that when there is microstructure noise, our proposed portfolios based on CLIME using either 5 min returns or pre-averaged 1 min returns can track the optimal portfolio closely, with risks close to the minimum risk. The plug-in portfolio carries a much higher risk.

Next, we again use $\widehat{R}_{\text{CLIME}}$ to estimate the daily minimum risks with high-frequency data. In this case, we consider the estimation with either RCV using 5-min data or PAV using 1-min data. The resulting estimated annualized risks are 8.73% and 8.82%, respectively. On the other hand, the estimate based on (3.5) using daily returns is 8.76%. We see that all the three estimates are similar to each other, and are roughly at the same level as the realized risk of the infeasible oracle portfolio.

5 Empirical Studies

In this section, we conduct empirical studies and compare minimum variance portfolios constructed via different estimators as well as the equally-weighted portfolio. We consider the NYSE composites of SP100 Index during 2011-2013. Among them, there are totally 80 stocks that remained as the composites of the Index and have complete data during this period. Our analysis is based on these 80 stocks. The observations are 1-min intra-day prices (390 observations for each stock per day, from 9:30:00 to 16:00:00). There are a total of $T = 251,\ 247,\ 249$ trading days for the year 2011, 2012 and 2013, respectively.

We include the following portfolios in our comparison.

(i) 1-Min-PAV-CLIME-based portfolio. This is based on Theorem 2 with pre-averaging estimator using 1-min returns with daily re-balancing. To be more specific, for each day t, we first compute $\widehat{\boldsymbol{\Sigma}}_{PAV,t}$ using 1-min data, and then construct an estimate of $\boldsymbol{\Omega}_t$ by replacing $\widehat{\boldsymbol{\Sigma}}$ with the $\widehat{\boldsymbol{\Sigma}}_{PAV,t}$ in (2.2). We then build a portfolio for day $t+1$:

$$\boldsymbol{w}_{t+1} = \frac{\widehat{\boldsymbol{\Omega}}_t \mathbf{1}}{\mathbf{1}^\intercal \widehat{\boldsymbol{\Omega}}_t \mathbf{1}}, \quad t = 0, 1, \ldots, T-1;$$

(ii) Daily CLIME-based portfolio. This portfolio weights are recalculated monthly: at the beginning of each month, we compute RCV based on daily returns during the immediate past 12 months, and then construct the portfolio using CLIME;

(iii) FFL portfolio. This is based on Fan et al. (2008). We use the S&P 500 ETF (SPY) as the factor. For each month, we use the daily returns during the past 12 months to construct the covariance matrix estimator $\widehat{\boldsymbol{\Sigma}}_t$, and plug its inverse into equation (2.3);

(iv) Nonlinear shrinkage portfolio. This is based on Ledoit and Wolf (2012). For each month, we use the daily returns during the past 12 months to construct the nonlinear shrinkage estimator $\widehat{\boldsymbol{\Omega}}_t$, and plug it into equation (2.3);

(v) Equally-weighted portfolio. This serves as the benchmark portfolio for comparison.

We take $S_0 = \$100$ for all portfolios to be compared. In addition to the standard deviation as a measure of risk, we consider the maximum drawdown (MDD) of each portfolio. The MDD measures the largest loss from a peak to a trough of a portfolio before a new peak is reached. Specifically,

$$\text{MDD} = \min_{\tau \in (0,T)} \frac{\big(P(\tau) - \max_{t\in(0,\tau)} P(t)\big)}{\max_{t\in(0,\tau)} P(t)}, \tag{5.1}$$

where $P(t)$ represents the portfolio value at day t.

Year 2012:

Figure 4 shows the time series plot of five portfolio values in Year 2012:

[*Fig* 4 *here*]

Table 3 reports some summary statistics for these portfolios.

[*Table* 3 *here*]

From Table 3, we observe that 1-Min-PAV-CLIME-based portfolio carries a low risk and a low maximum drawdown, and at the same time, it yields a high compound return.

Finally, the minimum risk estimated by (3.2) and (3.5) with daily returns, or (3.3) using 1-min intra-day returns are 5.54%, 5.81% and 5.45%, respectively. Comparing such estimates with Table 3 suggests that CLIME-based portfolios nearly achieve the minimum risk.

Year 2013:

Figure 5 shows the time series plot of the five portfolio values in Year 2013:

$[Fig\ 5\ here]$

Some summary statistics for these portfolios are reported in Table 4.

$[Table\ 4\ here]$

The minimum risk estimates, for Year 2013, based on (3.2) and (3.5) using daily returns or (3.3) using 1-min intra-day returns are 4.92%, 6.12% and 5.16%, respectively. These numbers again suggest that CLIME-based portfolios nearly achieve the minimum risk with low MDD and high compound returns.

Summarizing the comparisons in Year 2012 and 2013, we conclude the following:

(i) For both years tested, based on daily returns, the CLIME-based portfolios have lower risks than the equally-weighted, FFL and nonlinear shrinkage portfolios.

(ii) Using high-frequency data, specifically, one-minute returns in our case, leads to further improvements. The 1-Min-PAV-CLIME-based portfolios have similar risks to CLIME-daily, yet may yield (substantially) higher compound returns.

(iii) The empirical risks of all portfolios studied above are higher than the estimated minimum risks. The risks of CLIME-based portfolios are relatively closer to the estimated minimum risk, suggesting the gain of our approach.

6 Conclusion and Discussions

We propose in this paper estimators of the minimum variance portfolio in the high-dimensional setting based on high-frequency data. The desired risk convergence result (1.8) for the estimated portfolio is obtained under certain sparsity assumptions on the precision matrix. We further propose consistent estimators of the minimum risk, one based on high-frequency data under the same sparsity assumptions and the other based on low-frequency data without assuming sparsity.

Numerical studies demonstrate that our methods perform favorably. An important observation is that high-frequency volatility/covolatility estimation techniques can add value to portfolio allocation.

In this paper, we do not incorporate factor structure in our model. Empirical studies show that our methods compare favorably with the FFL portfolio that makes explicit use of the factor structure.

Acknowledgements

We thank the participants of SoFiE 2016 and FERM 2016 conferences for helpful discussions on preliminary versions of this paper. We are grateful to constructive comments from Robert Engle, Raymond Kan and Olivier Ledoit. Research partially supported by NSF grant DMS-1712735 and NIH grant R01 GM-123065; and the RGC grants GRF16305315, GRF16304317, GRF 16502014 and GRF 16518716 of the HKSAR.

REFERENCES

Ao, M., Li, Y., and Zheng, X. (2017), "Solving the Markowitz Optimization Problem for Large Portfolios," *working paper*.

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, 1481–1536.

— (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.

Basak, G. K., Jagannathan, R., and Ma, T. (2009), "Jackknife Estimator for Tracking Error Variance of Optimal Portfolios," *Management Science*, 55, 990–1002.

Cai, T. and Liu, W. (2011), "A direct estimation approach to sparse linear discriminant analysis," *Journal of the American Statistical Association*, 106, 1566–1577.

Cai, T. T., Liu, W., and Luo, X. (2011), "A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation," *Journal of the American Statistical Association*, 106, 594–607.

Chan, L. K., Karceski, J., and Lakonishok, J. (1999), "On portfolio optimization: Forecasting covariances and choosing the risk model," *Review of Financial Studies*, 12, 937–974.

Christoffersen, P. (2012), *Elements of Financial Risk Management*, Oxford: Academic Press.

Clarke, R. G., De Silva, H., and Thorley, S. (2006), “Minimum-variance portfolios in the US equity market,” *The Journal of Portfolio Management*, 33, 10–24.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009), “A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms,” *Management Science*, 55, 798–812.

Ding, Y., Li, Y., and Zheng, X. (2017), “Estimating High Dimensional Minimum Variance Portfolio under Factor Model,” *Working paper*.

El Karoui, N. (2010), “High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: risk underestimation,” *The Annals of Statistics*, 38, 3487–3566.

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

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, 412–428.

Fan, J., Zhang, J., and Yu, K. (2012b), “Vast portfolio selection with gross-exposure constraints,” *Journal of the American Statistical Association*, 107, 592–606.

Gloter, A. and Jacod, J. (2001), “Diffusions with measurement errors. I. Local asymptotic normality,” *ESAIM: Probability and Statistics*, 5, 225–242.

Haugen, R. A. and Baker, N. L. (1991), “The efficient market inefficiency of capitalization-weighted stock portfolios,” *The Journal of Portfolio Management*, 17, 35–40.

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, 2249–2276.

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

Kim, D. and Wang, Y. (2016), "Sparse PCA Based on High-Dimensional Ito processes with Measurement Errors," *Journal of Multivariate Analysis*, 152, 172–189.

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

Markowitz, H. (1952), "Portfolio Selection," *The Journal of Finance*, 7, 77–91.

Merton, R. C. (1980), "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 8, 323–361.

Podolskij, M. and Vetter, M. (2009), "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15, 634–658.

Schwartz, T. (2000), "How to Beat the S&P 500 with Portfolio Optimization," *working paper*.

Xia, N. and Zheng, X. (2017), "On the Inference about the Spectral Distribution of High-Dimensional Covariance Matrix Based on High-Frequency Noisy Observations," *Annals of Statistics, forthcoming*.

Xiu, D. (2010), "Quasi-maximum likelihood estimation of volatility with high frequency data," *Journal of Econometrics*, 159, 235–250.

Zhang, L. (2006), “Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach,” *Bernoulli*, 12, 1019–1043.

— (2010), “Estimating Covariation: Epps Effect, Microstructure Noise,” *Journal of Econometrics*, 160, 33–47.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005), “A tale of two time scales: determining integrated volatility with noisy high-frequency data,” *Journal of the American Statistical Association*, 100, 1394–1411.

Zheng, X. and Li, Y. (2011), “On the estimation of integrated covariance matrices of high dimensional diffusion processes,” *Annals of Statistics*, 39, 3121–3151.

A Appendix: Proofs

We start with a result about the CLIME estimator. It is a direct consequence of Theorem 6 in Cai and Liu (2011) .

Proposition 1. *Suppose that* $\mathbf{\Omega} \in \mathcal{U}(q, s_0, M)$. *If* $\lambda \geq M||\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}||_\infty$ *with* $\widehat{\mathbf{\Sigma}}$ *the estimator for* $\mathbf{\Sigma}$ *used to construct CLIME estimator, then we have*

$$||\widehat{\mathbf{\Omega}} - \mathbf{\Omega}||_2 \leq C s_0 M^{1-q} \lambda^{1-q}, \tag{A.1}$$

where $C \leq 2(1 + 2^{1-q} + 3^{1-q})4^{1-q}$.

Next, we show that if (A.1) holds, then under Assumption A and if $s_0 M^{1-q}\lambda^{1-q} \to 0$, for all p large enough, we have

$$\left| \frac{\mathbf{1}^\intercal \mathbf{\Omega} \mathbf{1}}{\mathbf{1}^\intercal \widehat{\mathbf{\Omega}} \mathbf{1}} - 1 \right| \leq \frac{2}{\delta} ||\widehat{\mathbf{\Omega}} - \mathbf{\Omega}||_2. \tag{A.2}$$

In fact, if $s_0 M^{1-q}\lambda^{1-q} \to 0$, then (A.1) implies that $||\widehat{\mathbf{\Omega}} - \mathbf{\Omega}||_2 \leq \delta/2$ for all p large enough, in which case we have

$$\left| \frac{\mathbf{1}^\intercal \mathbf{\Omega} \mathbf{1}}{\mathbf{1}^\intercal \widehat{\mathbf{\Omega}} \mathbf{1}} - 1 \right| \leq \frac{p||\widehat{\mathbf{\Omega}} - \mathbf{\Omega}||_2}{|\mathbf{1}^\intercal \mathbf{\Omega} \mathbf{1}| - p||\widehat{\mathbf{\Omega}} - \mathbf{\Omega}||_2} \leq \frac{p||\widehat{\mathbf{\Omega}} - \mathbf{\Omega}||_2}{p(1/\lambda_{\max} - \delta/2)} \leq \frac{2}{\delta} ||\widehat{\mathbf{\Omega}} - \mathbf{\Omega}||_2.$$

Moreover, for the risk R_n in (2.4) of the portfolio (2.3), we have

$$\begin{aligned} \frac{R_n}{R_{\min}} &= \frac{(\widehat{\mathbf{\Omega}}\mathbf{1})^\intercal \mathbf{\Sigma}(\widehat{\mathbf{\Omega}}\mathbf{1})}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}} \left(\frac{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}}\right) \\ &= \left(1 + \frac{(\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{\Omega}\mathbf{1})^\intercal \mathbf{\Sigma}(\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{\Omega}\mathbf{1})}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}} + \frac{\mathbf{1}^\intercal(\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{\Omega}\mathbf{1})}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}}\right)\left(\frac{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}}\right) \\ &= \left(1 + \frac{(\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{\Omega}\mathbf{1})^\intercal \mathbf{\Sigma}(\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{\Omega}\mathbf{1})}{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}\frac{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}} + \frac{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}\frac{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}}\right)\left(\frac{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}}\right). \end{aligned}$$

Because

$$|\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{\Omega}\mathbf{1}|_2 \le \sqrt{p}||\widehat{\mathbf{\Omega}}-\mathbf{\Omega}||_2,$$

$$|\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}| \le p||\widehat{\mathbf{\Omega}}-\mathbf{\Omega}||_2,$$

we have

$$\begin{aligned} \left|\frac{(\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{\Omega}\mathbf{1})^\intercal \mathbf{\Sigma}(\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{\Omega}\mathbf{1})}{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}\right| &\le \frac{\lambda_{max}|\widehat{\mathbf{\Omega}}\mathbf{1}-\mathbf{\Omega}\mathbf{1}|_2^2}{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}} \\ &\le \frac{p\lambda_{max}^2}{p}||\widehat{\mathbf{\Omega}}-\mathbf{\Omega}||_2^2 \\ &\le \frac{1}{\delta^2}||\widehat{\mathbf{\Omega}}-\mathbf{\Omega}||_2^2. \end{aligned}$$

Therefore, by (A.2) and (A.1),

$$\begin{aligned} \frac{R_n}{R_{\min}} &= \left(1+O(||\widehat{\mathbf{\Omega}}-\mathbf{\Omega}||_2)\right)\left(1+O(||\widehat{\mathbf{\Omega}}-\mathbf{\Omega}||_2)\right) \\ &= \left(1+O(||\widehat{\mathbf{\Omega}}-\mathbf{\Omega}||_2)\right) \\ &= \left(1+O\left(s_0 M^{1-q}\lambda^{1-q}\right)\right). \end{aligned} \tag{A.3}$$

A.1 Proof of Theorem 1

Proof. To prove the theorem, it is sufficient to show that under the Assumption A-E, we have for $\eta > 0$,

$$P\left(||\widehat{\mathbf{\Sigma}}_{RCV} - \mathbf{\Sigma}_{ICV}||_\infty > \eta\sqrt{\log p/n}\right) \leq \frac{C_0}{p^{C_1\eta^2-2}}, \tag{A.4}$$

for some positive constants C_0, C_1. Then with $\lambda = \eta M\sqrt{\log p/n}$, we have, from Proposition 1 and (A.3), that with probability greater than $1 - \frac{C_0}{p^{C_1\eta^2-2}}$,

$$\frac{R_n}{R_{\min}} - 1 = O\left(M^{2-2q}s_0\left((\log p/n)^{(1-q)/2}\right)\right),$$

where R_n is the risk associated with the portfolio $\widehat{\boldsymbol{w}} = \frac{\widehat{\mathbf{\Omega}}^{RCV}\mathbf{1}}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}^{RCV}\mathbf{1}}$, and $\widehat{\mathbf{\Omega}}^{RCV}$ is the CLIME estimator obtained by replacing $\widehat{\mathbf{\Sigma}}$ with $\widehat{\mathbf{\Sigma}}_{RCV}$ in (2.2).

We now prove (A.4). We first consider the estimation of integrated volatility and co-volatility.

Lemma 1. *Suppose that* (X_t) *satisfies*

$$dX_t = \mu_t\ dt + \sigma_t\ dW_t, \quad t \in [0,1],$$

and there exist constants C_μ, C_σ *such that*

$$|\mu_t| \leq C_\mu, \quad \textit{and} \quad |\sigma_t| \leq C_\sigma < \infty, \quad \textit{for all } t \in [0,1] \quad \textit{almost surely.}$$

Suppose further that the observation times t_i^n *satisfy (2.9) in Assumption D. Denote the* *realized volatility as* $[X, X]_t := \sum_{\{i: t_i^n \le t\}} (X_{t_i^n} - X_{t_{i-1}^n})^2$*. Then, we have*

$$P\left(\sqrt{n}|[X,X]_1 - \int_0^1 \sigma_t^2 \, dt| > x\right) \le C \exp\left(-x^2/(32 C_\sigma^4 C_\Delta^2)\right) \tag{A.5}$$

for all $0 \le x \le 2C_\sigma^2 C_\Delta \sqrt{n}$*, where the constant* $C > 0$ *depends on only* C_σ *and* C_Δ *and can be specified.*

Proof. For notational ease, we shall omit the superscript n and write t_i for t_i^n etc. Define $\widetilde{X}_t = X_0 + \int_0^t \sigma_s dW_s$. Then we have $X_t = \widetilde{X}_t + \int_0^t \mu_s ds$. Observe that

$$\begin{aligned}
&[X,X]_t - \int_0^t \sigma_s^2 \, ds \\
=&\left([\widetilde{X},\widetilde{X}]_t - \int_0^t \sigma_s^2 \, ds\right) + \sum_{t_i \le t} \left(\int_{t_{i-1}}^{t_i} \mu_s ds\right)^2 + 2\sum_{t_i \le t}\left(\widetilde{X}_{t_i} - \widetilde{X}_{t_{i-1}}\right) \cdot \int_{t_{i-1}}^{t_i} \mu_s ds \\
:=& I_1 + I_2 + I_3.
\end{aligned}$$

We first consider terms I_1 and I_2. By Eqn.(A.6) in Fan et al. (2012a), we have for $|\theta| \le \sqrt{n}/(4C_\sigma^2 C_\Delta)$,

$$E\left(\exp(\theta\sqrt{n} \cdot I_1)\right) \le \exp(2\theta^2 C_\sigma^4 C_\Delta^2). \tag{A.6}$$

As to term I_2, it is easy to see that $I_2 \le C_\mu^2 C_\Delta / n$. It follows that for $|\theta| \le \sqrt{n}/(4C_\sigma^2 C_\Delta)$,

$$E\left(\exp(\theta\sqrt{n}(I_1 + I_2))\right) \le \quad C_1 \exp(2\theta^2 C_\sigma^4 C_\Delta^2),$$

where $C_1 = \exp(C_\mu^2/(4C_\sigma^2))$. By Markov's inequality, we then obtain that for $0 \le \theta \le$

$\sqrt{n}/(4C_\sigma^2 C_\Delta)$,

$$P(\sqrt{n}|I_1+I_2|>x) \le \exp(-\theta x) E\left(\exp\{\theta\sqrt{n}|I_1+I_2|\}\right)$$
$$\le 2C_1 \exp\left(2\theta^2 C_\sigma^4 C_\Delta^2 - \ x\ \theta\right).$$

Taking $\theta = x/4C_\sigma^4 C_\Delta^2$ yields that for $x \in [0, \sqrt{n}C_\sigma^2 C_\Delta]$,

$$P(\sqrt{n}|I_1+I_2|>x) \le 2C_1 \exp\left(-\frac{x^2}{8C_\sigma^4 C_\Delta^2}\right). \tag{A.7}$$

Next we consider term I_3. By Cauchy-Schwartz inequality, we have

$$|I_3| \le 2\sqrt{[\widetilde{X},\widetilde{X}]_1 I_2} \le \frac{2C_\mu\sqrt{C_\Delta}}{\sqrt{n}}\sqrt{[\widetilde{X},\widetilde{X}]_1} \le \frac{2C_\mu C_\Delta}{\sqrt{n}}\sqrt{[\widetilde{X},\widetilde{X}]_1}.$$

Therefore

$$P(\sqrt{n}|I_3|>x) \le P(4C_\mu^2 C_\Delta^2 [\widetilde{X},\widetilde{X}]_1 > x^2)$$
$$\le E\exp\left(\frac{C_\mu^2[\widetilde{X},\widetilde{X}]_1}{2C_\sigma^4} - \frac{x^2}{8C_\sigma^4 C_\Delta^2}\right).$$

By (A.6) and the assumption that $|\sigma_t| \le C_\sigma$, we obtain that when $n \ge 2C_\mu^2 C_\Delta^2/C_\sigma^2$,

$$\begin{aligned} P(\sqrt{n}|I_3|>x) &\le \exp\left(\frac{C_\mu^4 C_\Delta^2}{2C_\sigma^4\ n} + \frac{C_\mu^2}{2C_\sigma^2} - \frac{x^2}{8C_\sigma^4 C_\Delta^2}\right) \\ &\le C_2 \exp\left(-\frac{x^2}{8C_\sigma^4 C_\Delta^2}\right), \end{aligned} \tag{A.8}$$

where $C_2 = \exp(C_\mu^2/C_\sigma^2)$.

Combining (A.7) and (A.8) we see that for all $n \ge 2C_\mu^2 C_\Delta^2/C_\sigma^2$ and all $0 \le x \le$

$2C_\sigma^2 C_\Delta \sqrt{n}$,

$$\begin{aligned}
&P(\sqrt{n}|[X,X]_1 - \int_0^1 \sigma_t^2 \, dt| > x) \\
\le &P(\sqrt{n}|I_1 + I_2| > x/2) + P(\sqrt{n}|I_3| > x/2) \\
\le &(2C_1 + C_2) \exp\left(-\frac{x^2}{32 C_\sigma^4 C_\Delta^2} \right).
\end{aligned}$$

The conclusion in the lemma follows by taking $C = 2C_1 + C_2$. □

Now we focus on estimating co-volatility between two log-price processes. We assume that two log price processes X and Y satisfy

$$dX_t = X_0 + \mu_t^X dt + \sigma_t^X dW_t^X \quad \text{and} \quad dY_t = Y_0 + \mu_t^Y dt + \sigma_t^Y dW_t^Y, \tag{A.9}$$

where $\mathrm{Corr}(W_t^X, W_t^Y) = \rho_t^{(X,Y)}$.

Lemma 2. *Suppose that the two processes (X, Y) satisfy (A.9). Furthermore, assume that there exist finite positive constants C_μ, C_σ and C_Δ such that $|\mu_t^i| \le C_\mu < \infty$, $|\sigma_t^i| \le C_\sigma < \infty, a.s.$, for $i = X, Y$. Suppose also that the two processes are observed at times $\{t_n^i\}$ which satisfy (2.9) in Assumption D. Then, we have for all $0 \le x \le 4C_\sigma^2 C_\Delta \sqrt{n}$,*

$$P\left(\sqrt{n}\big|[X,Y]_1 - \int_0^1 \sigma_t^X \sigma_t^Y \rho_t^{(X,Y)} dt\big| > x \right) \le 2C \exp\left(-\frac{x^2}{128 C_\sigma^4 C_\Delta^2} \right),$$

where the constant $C > 0$ depends on only C_σ and C_Δ and can be specified.

Proof. Again we shall omit the superscript n and write t_i for t_i^n etc. Define $Z^\pm = X \pm Y$. Then

$$[X,Y]_1 = \frac{1}{4}([Z^+, Z^+]_1 - [Z^-, Z^-]_1). \tag{A.10}$$

Also both $Z^{\pm}$ are stochastic processes with bounded drifts and volatilities. As a matter of fact, they can be represented by

$$dZ_t^{\pm} = \mu_{Z_t^{\pm}} dt + \sigma_t^{Z^{\pm}} dW_t^{\pm},$$

where $\mu_{Z_t^{\pm}} = \mu_t^X \pm \mu_t^Y$, $\sigma_t^{Z^{\pm}} = \sqrt{(\sigma_t^X)^2 + (\sigma_t^Y)^2 \pm 2\rho_t^{(X,Y)} \sigma_t^X \sigma_t^Y}$, and $W^{\pm}$ are standard Brownian motions. Therefore, they both satisfy the conditions in Lemma 1 with bound on the drift terms being $2C_\mu$ and bound on the volatility terms being $2C_\sigma$. Thus, by Lemma 1, we have that for all $0 \le x \le 8C_\sigma^2 C_\Delta \sqrt{n}$

$$P\left(\sqrt{n}|[Z^{\pm}, Z^{\pm}]_1 - \int_0^1 (\sigma_t^{Z^{\pm}})^2 \; dt| > x\right) \le C \exp\left(-x^2/(512 C_\sigma^4 C_\Delta^2)\right).$$

It follows from the decomposition (A.10) that

$$P\left(\sqrt{n}\big|[X,Y]_1 - \int_0^1 \sigma_t^X \sigma_t^Y \rho_t^{(X,Y)} dt\big| > x\right) \le 2C \exp\left(-x^2/(128 C_\sigma^4 C_\Delta^2)\right).$$

□

With the previous two lemmas, we obtain that there exist positive constants C_0, C_1 such that for $0 \le x \le 4C_\sigma^2 C_\Delta \sqrt{n}$,

$$\max_{i,j} P(\sqrt{n}|\hat{\sigma}_{ij} - \sigma_{ij}| > x) \le C_0 \exp(-C_1 x^2),$$

where $(\sigma_{ij}) =: \mathbf{\Sigma}_{ICV}$ and $(\hat{\sigma}_{ij}) =: \widehat{\mathbf{\Sigma}}_{RCV}$. Therefore, by the Bonferroni inequality, we have

$$\begin{aligned} P\left(||\widehat{\mathbf{\Sigma}}_{RCV} - \mathbf{\Sigma}_{ICV}||_\infty > \eta\sqrt{\log p/n}\right) &\leq \sum_{1\leq i,j\leq p} P(|\hat{\sigma}_{ij} - \sigma_{ij}| > \eta\sqrt{\log p/n}) \\ &\leq p^2 \cdot C_0 \exp(-C_1\eta^2 \log(p)) = \frac{C_0}{p^{C_1\eta^2-2}}. \end{aligned}$$

This finishes the proof.

□

A.2 Proof of Theorem 2

To prove Theorem 2, similar to the proof of Theorem 1, it is sufficient to provide a bound on the the element-wise estimation error in using $\widehat{\mathbf{\Sigma}}_{PAV}$ to estimate $\mathbf{\Sigma}$. The following result from Kim and Wang (2016) gives an exponential tail bound for this element-wise estimation error.

Proposition 2. *[Theorem 1 in Kim and Wang (2016)] Suppose that* $(\mathbf{X}_t)$ *satisfies* (2.1). *Under Assumption A-D and F-H, the PAV estimator* $\widehat{\mathbf{\Sigma}}_{PAV}$ *satisfies that*

$$P\left(|(\widehat{\mathbf{\Sigma}}_{PAV} - \mathbf{\Sigma}_{ICV})_{ij}| \geq x\right) \leq C_2 \exp(-\sqrt{n}C_3x^2),$$

where $(\widehat{\mathbf{\Sigma}}_{PAV} - \mathbf{\Sigma}_{ICV})_{ij}$ *denotes the* ij*-th entry of the matrix* $(\widehat{\mathbf{\Sigma}}_{PAV} - \mathbf{\Sigma}_{ICV})$*, and* x *is a positive number in a neighbor of 0, and* C_2 *and* C_3 *are positive constants independent of* n *and* p*.*

It follows from the Bonferroni inequality that

$$
\begin{aligned}
P\left(||\widehat{\mathbf{\Sigma}}_{PAV} - \mathbf{\Sigma}_{ICV}||_\infty \geq \eta\sqrt{\log p}/n^{1/4}\right) &\leq \sum_{1\leq i,j\leq p} P\left(|(\widehat{\mathbf{\Sigma}}_{PAV} - \mathbf{\Sigma})_{ij}| \geq \eta\sqrt{\log p}/n^{1/4}\right) \\
&\leq \sum_{1\leq i,j\leq p} C_2 \exp(-\sqrt{n}C_3(\eta\sqrt{\log p}/n^{1/4})^2) \\
&= p^2 C_2 \exp(-(C_3\eta^2)\log p) \\
&= \frac{C_2}{p^{(C_3\eta^2-2)}}.
\end{aligned} \tag{A.11}
$$

Therefore, with $\lambda = \eta M\sqrt{\log p}/n^{1/4}$, we have, from (A.3), that with probability greater than $1 - \frac{C_2}{p^{(C_3\eta^2-2)}}$,

$$
\frac{R_n}{R_{\min}} - 1 = O\left(M^{2-2q}s_0\left((\log p)^{(1-q)/2}/n^{(1-q)/4}\right)\right),
$$

where R_n is the risk associated with the portfolio $\widehat{\boldsymbol{w}} = \frac{\widehat{\mathbf{\Omega}}^{PAV}\mathbf{1}}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}^{PAV}\mathbf{1}}$, and $\widehat{\mathbf{\Omega}}^{PAV}$ is the CLIME estimator obtained by replacing $\widehat{\mathbf{\Sigma}}$ with $\widehat{\mathbf{\Sigma}}_{PAV}$ in (2.2). This completes the proof.

A.3 Proof of Theorem 3

Proof. Note that $\frac{\widehat{R}_{\text{CLIME}}}{R_{\min}} = \frac{\mathbf{1}^\intercal\mathbf{\Omega}\mathbf{1}}{\mathbf{1}^\intercal\widehat{\mathbf{\Omega}}\mathbf{1}}$ where $\widehat{\mathbf{\Omega}}$ is the CLIME estimator based on RCV in the noiseless case and PAV in the noisy case. The result for the noiseless case follows directly from (A.2) and (A.4), and in the noisy case follows from (A.2) and (A.11). □

A.4 Proof of Theorem 4

Proof. The first result is a direct consequence of Theorem 4.1 in El Karoui (2010), by which we have

$$\frac{\mathbf{1}^\intercal \mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^\intercal \mathbf{\Sigma}^{-1}\mathbf{1}} \to_p \frac{1}{(1-\rho)}, \tag{A.12}$$

The conclusion in the theorem follows by noticing that

$$\frac{\widehat{R}_p}{R_{\min}} = \frac{\boldsymbol{w}_p^\intercal \mathbf{S} \boldsymbol{w}_p}{\boldsymbol{w}_{\text{opt}}^\intercal \mathbf{\Sigma} \boldsymbol{w}_{\text{opt}}} = \frac{\mathbf{1}^\intercal \mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\intercal \mathbf{S}^{-1}\mathbf{1}}.$$

For the CLT result, by Theorem 3.1 in El Karoui (2010), we have

$$(n-1)\frac{\widehat{R}_p}{R_{\min}} \sim \chi^2_{n-p}, \tag{A.13}$$

where $\widehat{R}_p = \boldsymbol{w}_p^\intercal \mathbf{S} \boldsymbol{w}_p$ is the perceived risk of "plug-in" portfolio. Therefore,

$$(1-\rho)(n-1)\frac{\widehat{R}_{\min}}{R_{\min}} = \frac{(n-p)(n-1)}{n}\frac{\widehat{R}_{\min}}{R_{\min}} \sim \chi^2_{n-p}.$$

It follows that, as n, p and $n - p \to \infty$,

$$\sqrt{n-p}\left(\frac{n-1}{n}\frac{\widehat{R}_{\min}}{R_{\min}} - 1\right) \Rightarrow N(0,2).$$

The convergence (3.7) follows. □

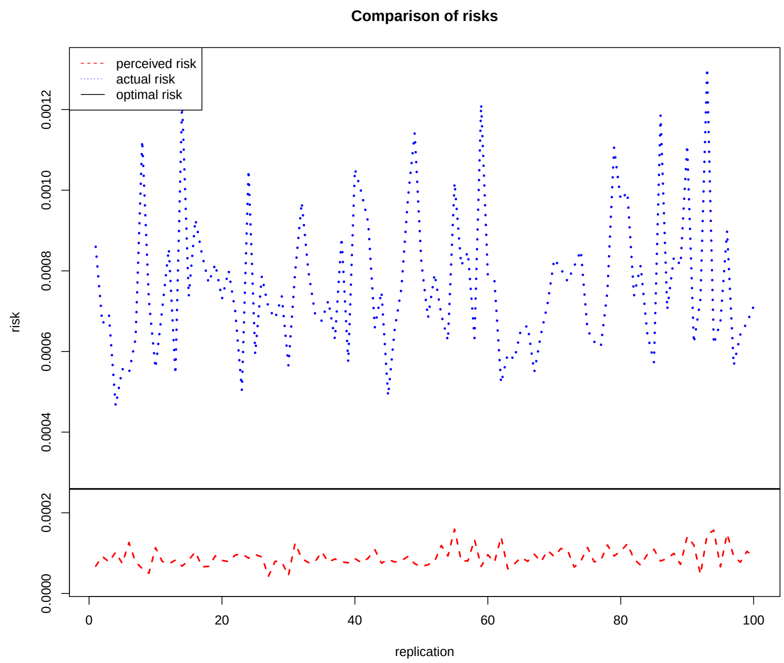

Figure 1. Comparison of actual and perceived risks of the plug-in portfolio. The portfolios are constructed based on returns simulated from i.i.d. multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma}^{1/2}$ where $\mathbf{\Sigma} = 0.05(0.7^{|i-j|})_{i,j=1,\ldots,p}$. The number of assets and observations are 50 and 75, respectively. The comparison is replicated 100 times.

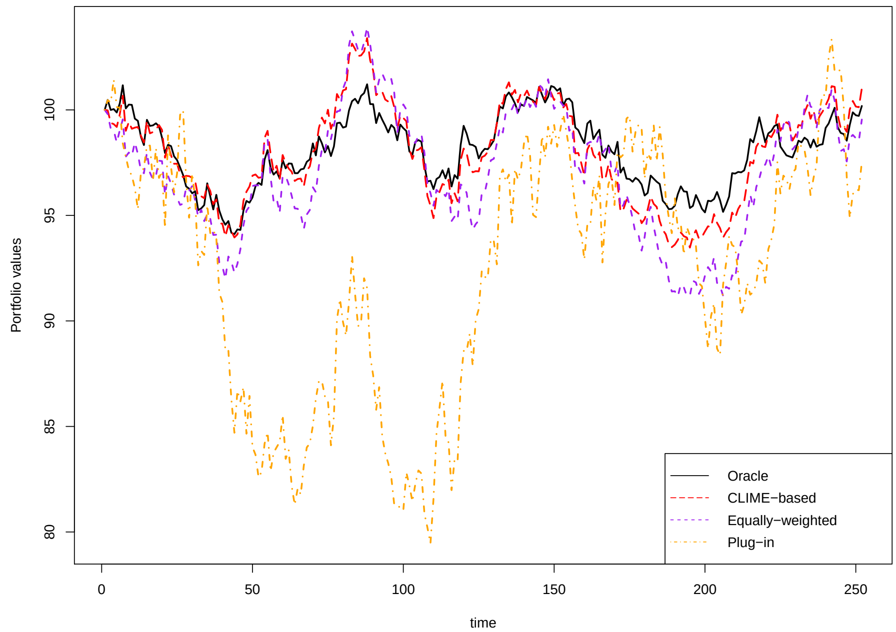

Figure 2. Time series plots of portfolio values. The portfolios are built based on the simulated prices with no microstructure noise.

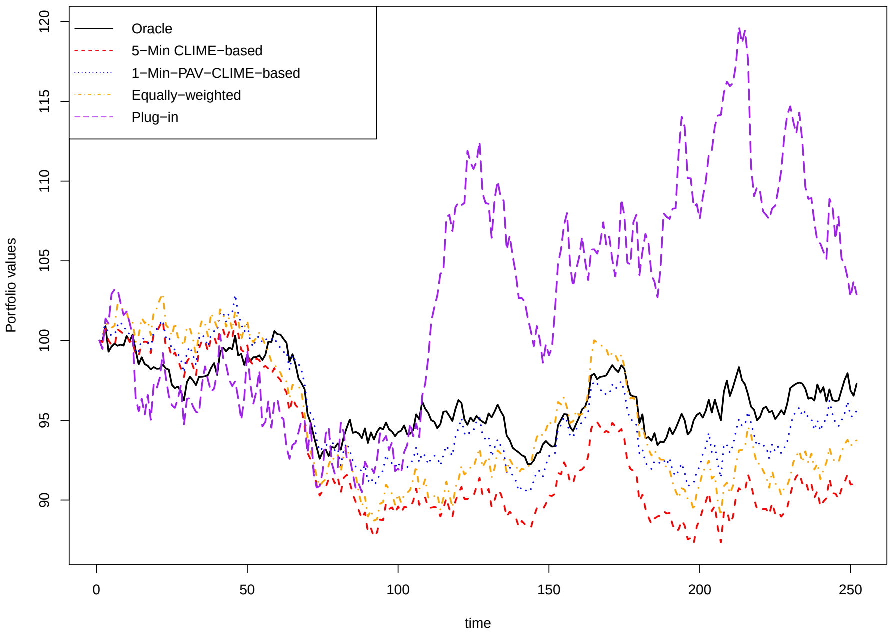

Figure 3. Time series plots of portfolio values. The portfolios are built based on the simulated prices with microstructure noise.

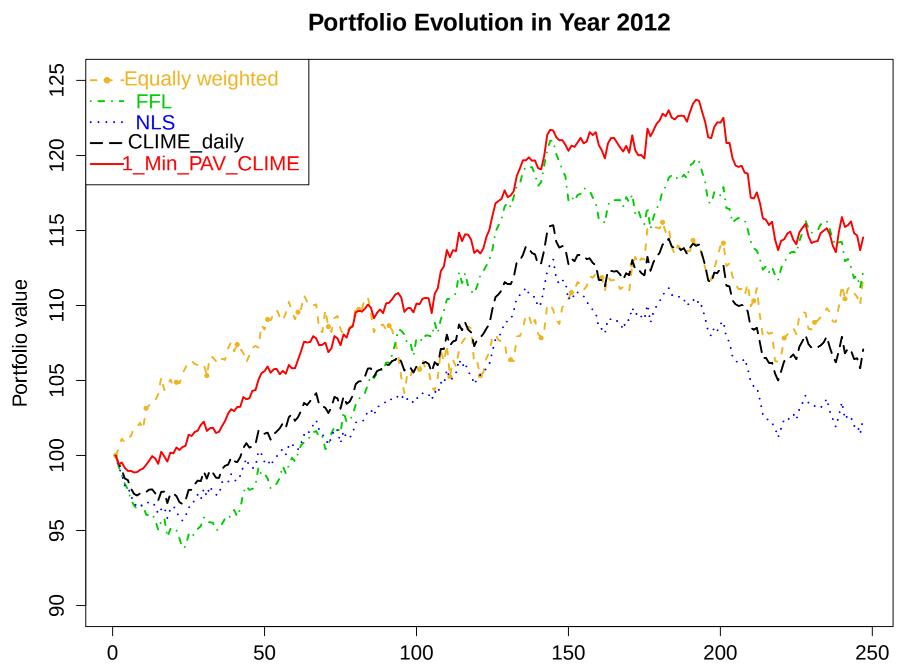

Figure 4. Evolution of different portfolios under comparison based on the NYSE stocks in Year 2012. All portfolios start with value $100.

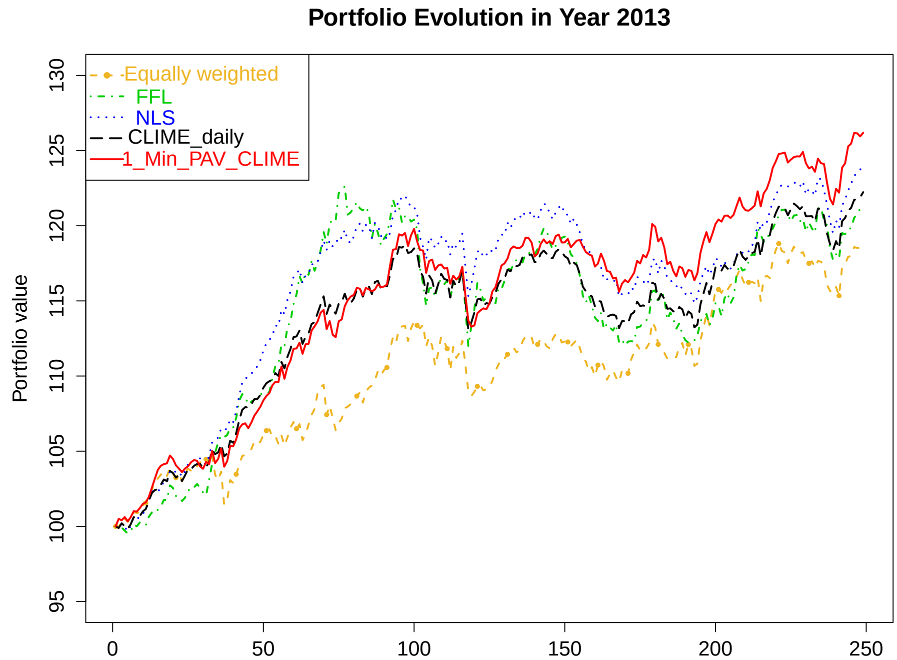

Figure 5. Evolution of different portfolios under comparison based on the NYSE stocks in Year 2013. All portfolios start with value $100.

Portfolios	Oracle	CLIME-based	Equally-weighted	Plug-in
Annualized SD	8.0%	9.33%	11.71%	25.42%

Table 1.
Annualized standard deviations of the portfolios under comparison based on the simulated prices with no microstructure noise.

Portfolios	Annualized SD of daily return
Oracle	9.2%
5-Min CLIME-based	10.7%
1-Min-PAV-CLIME-based	10.4%
Equally-weighted	12.1%
Plug-in	22.9%

Table 2.
Annualized standard deviations of the portfolios under comparison based on the simulated prices with microstructure noise.

Portfolio	Ann. SD of daily return	Compound return	MDD
Equally-weighted	10.45%	11.97%	−8.04%
FFL	8.15%	12.26%	−8.12%
Nonlinear shrinkage	7.42%	2.52%	−10.49%
CLIME-daily	7.16%	7.07%	−8.97%
1-Min-PAV-CLIME	7.23%	14.53%	−8.21%

Table 3. Comparison among the portfolios based on the NYSE data in Year 2012.

Portfolio	Ann. SD of daily return	Compound return	MDD
Equally-weighted	8.87%	18.89%	−4.37%
FFL	9.77%	21.15%	−8.67%
Nonlinear shrinkage	7.52%	24.05%	−5.87%
CLIME-daily	7.37%	22.22%	−4.75%
1-Min-PAV-CLIME	7.62%	26.19%	−5.43%

Table 4. Comparison among the portfolios based on the NYSE data in Year 2013.