Journal of Econometrics 214 (2020) 482–494

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

High-dimensional minimum variance portfolio estimation based on high-frequency data

T. Tony Cai [a], Jianchang Hu [b], Yingying Li [c,*], Xinghua Zheng [d]

[a] Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, United States of America
[b] Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, United States of America
[c] Department of ISOM and Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
[d] Department of ISOM, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 23 January 2018
Received in revised form 5 April 2019
Accepted 19 April 2019
Available online 7 September 2019

JEL classification:
C13, C55, C58, G11

Keywords:
Minimum variance portfolio
High dimension
High frequency
CLIME estimator
Precision matrix

ABSTRACT

This paper studies the estimation of high-dimensional minimum variance portfolio (MVP) based on the high frequency returns which can exhibit heteroscedasticity *and* possibly be contaminated by microstructure noise. Under certain sparsity assumptions on the precision matrix, we propose estimators of the MVP and prove that our portfolios asymptotically achieve 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 the favorable performance of the proposed portfolios.

1. Introduction

1.1. Background

Since the ground-breaking work of Markowitz (1952), the mean–variance portfolio has caught significant attention from both academics 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. Theoretically, mean–variance optimization has been criticized from the portfolio standpoint due to invalid preferences from the axioms of choice, inconsistent dynamic behavior, etc.

The minimum variance portfolio (MVP) has received growing attention over the past few years (see, e.g., DeMiguel et al. (2009a) and the references therein). It avoids the difficulties in estimating the expected returns and is on the efficient frontier. In addition, the MVP is 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 can enjoy

* Corresponding author.
E-mail addresses: tcai@wharton.upenn.edu (T.T. Cai), huj@stat.wisc.edu (J. Hu), yyli@ust.hk (Y. Li), xhzheng@ust.hk (X. Zheng).

https://doi.org/10.1016/j.jeconom.2019.04.039
0304-4076/

both lower risk and higher return 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. Over short time horizons, mean return is usually completely dominated by the volatility, consequently, as a prudent common practice, when the time horizon of interest is short, the expected returns are often assumed to be zero (see, e.g., 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.

In addition, there are benefits of using high-frequency data. On the one hand, large number of observations can potentially help facilitate better understanding of the covariance structure of returns. Developments in this direction in the high-dimensional setting include Wang and Zou (2010), Tao et al. (2011), Zheng and Li (2011), Tao et al. (2013), Kim et al. (2016), Aït-Sahalia and Xiu (2017), Xia and Zheng (2018), Dai et al. (2019), Pelger (2019), among others. On the other hand, high-frequency data allow 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 heteroscedasticity and microstructure noise, among others.

We consider in this paper the estimation of high-dimensional MVP using high-frequency data. To be more specific, given p assets, whose returns $\mathbf{X} = (X_1, \ldots, X_p)^\intercal$ have covariance matrix $\boldsymbol{\Sigma}$, we aim to find:

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

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

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

which yields the minimum risk

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

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

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

or its equivalent formulation:

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

Such a setting applies, for example, in leveraged investment. We remark 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 $\widetilde{\boldsymbol{\Sigma}} = \text{diag}(\beta_1, \ldots, \beta_p)\boldsymbol{\Sigma}\,\text{diag}(\beta_1, \ldots, \beta_p)$, 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 the paper, we focus on the problem (1.1).

A main challenge of solving the optimization problem (1.1) comes from high-dimensionality because modern portfolios often involve a large number of assets. See, for example, Zheng and Li (2011), Fan et al. (2012a), Fan et al. (2012b), Ao et al. (2018), Xia and Zheng (2018), and Dai et al. (2019) on issues about and progress made on vast portfolio management.

On the other hand, the estimation of the minimum risk defined in (1.3) is also a problem of interest. This provides a reference target for the estimated minimum variance portfolios.

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). Fig. 1 shows the risk of the plug-in portfolio based on 100 replications. One can see that the actual risk $R(\boldsymbol{w}_p) = \boldsymbol{w}_p^\intercal \boldsymbol{\Sigma} \boldsymbol{w}_p$ 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 lower 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?*

Because of such issues 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:

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

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 portfolio. 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

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

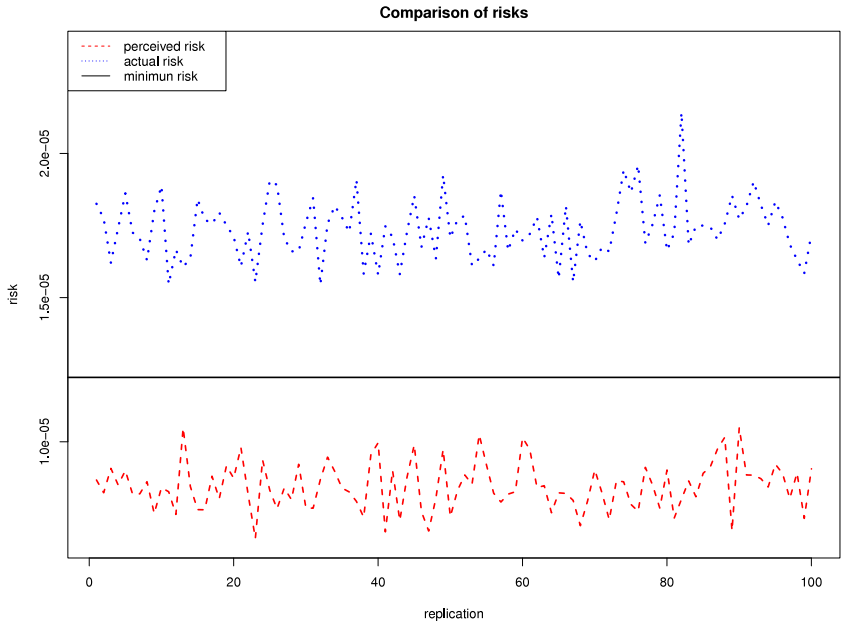

Fig. 1. Comparison of actual and perceived risks of the plug-in portfolio. The portfolios are constructed based on returns simulated from i.i.d. multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$ calibrated from real data; see Section 4 for details. The numbers of assets and observations are 80 and 252, respectively. The comparison is replicated 100 times.

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 integrated covariance matrix (see Section 2.1 for related background), and establish concentration inequalities for the element-wise estimation error. 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 (18) in Fan et al. (2012a) for the precise statement. In particular, bound (1.7) guarantees that under 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 sufficient to guarantee (near) optimality. In fact, under rather general assumptions (which do not exclude factor models), the minimum risk $R_{\min} = 1/\mathbf{1}^\intercal \boldsymbol{\Sigma}^{-1} \mathbf{1}$ may go to zero as the number of assets $p \to \infty$; see Ding et al. (2018) for a thorough discussion. If indeed the minimum risk goes to 0 as $p \to \infty$, then the difference $|R(\widehat{\boldsymbol{w}}_{\text{opt}}) - R_{\min}|$ going 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.,

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

where $\xrightarrow{p}$ stands for convergence in probability.

1.2. Main contributions of the paper

Our contributions mainly lie in the following aspects.

We propose estimators of minimum variance portfolio that can accommodate stochastic volatility and market microstructure noise, which are intrinsic to high-frequency returns. Under some sparsity assumptions on the inverse of the covariance matrix (also known as the precision matrix), our estimated portfolios enjoy the desired convergence (1.8).

We also introduce consistent estimators of the minimum risk. One such estimator does not depend on the sparsity assumption and also enjoys a CLT.

1.3. Organization of the paper

The paper is organized as follows. In Section 2, we present our estimators of the MVP and show that their risks converge to the minimum risk in the sense of (1.8). We have an estimator that incorporates stochastic volatility (CLIME-SV) and one

that incorporates stochastic volatility and microstructure noise (CLIME-SVMN). A consistent estimator of the minimum risk is proposed in Section 2.4, for which we also establish the CLT. An extension of our method to utilize factors is developed in Section 3. Section 4 presents simulation results to illustrate the performance of both portfolio and minimum risk estimations. Empirical study results based on S&P 100 Index constituents are reported in Section 5. We conclude our paper with a brief summary in Section 6. All proofs are given in Appendix.

2. Estimation methods and asymptotic properties

2.1. High-frequency data model

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

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

where $(\boldsymbol{\mu}_t) = (\mu_t^1, \ldots, \mu_t^p)^\intercal$ is the drift process, $(\boldsymbol{\Theta}_t) = (\theta_t^{ij})_{1\leq i,j\leq p}$ 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 $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$.

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 over an interval, say [0, 1], is

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

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

Let us emphasize that in general, $\boldsymbol{\Sigma}_{\text{ICV}}$ 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 of $\boldsymbol{\Sigma}_{\text{ICV}}$ based on historical data. The simplest approach is to assume that $\boldsymbol{\Sigma}_{\text{ICV},t} \approx \boldsymbol{\Sigma}_{\text{ICV},t+1}$,[1] where $\boldsymbol{\Sigma}_{\text{ICV},t}$ stands for the ICV matrix in period $[t-1, t]$. Under such an assumption, if we can construct a portfolio $\boldsymbol{w}$ based on the observations during $[t-1, t]$ (and hence only measurable to $\mathcal{F}_t$) that can approximately minimize the ex-post risk $\boldsymbol{w}^\intercal \boldsymbol{\Sigma}_{\text{ICV},t} \boldsymbol{w}$, then if we hold the portfolio during the next period $[t, t+1]$, the actual risk $\boldsymbol{w}^\intercal \boldsymbol{\Sigma}_{\text{ICV},t+1} \boldsymbol{w}$ is still approximately minimized. In this article we 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 original CLIME method is developed under the i.i.d. observation setting. Specifically, suppose one has n i.i.d. observations from a population with covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{\Omega} := \boldsymbol{\Sigma}^{-1}$ be the corresponding precision matrix. The CLIME estimator of $\boldsymbol{\Omega}$ is defined as

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

where $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix, $\mathbf{I}$ is the identity matrix, and for any matrix $A = (a_{ij})$, $\|A\|_1 := \sum_{i,j} |a_{ij}|$ and, recall that, $\|A\|_\infty = \max_{ij} |a_{ij}|$. The λ is a tuning parameter, and is usually chosen via cross-validation.

The CLIME method is designed for the following uniformity class of precision matrices. For any $0 \leq q < 1$, $s_0 = s_0(p) < \infty$ and $M = M(p) < \infty$, let

$$\mathcal{U}(q, s_0, M) = \left\{ \boldsymbol{\Omega} = (\boldsymbol{\Omega}_{ij})_{p\times p} : \boldsymbol{\Omega} \text{ positive definite}, \ \|\boldsymbol{\Omega}\|_{L_1} \leq M, \max_{1\leq i\leq p} \sum_{j=1}^p |\boldsymbol{\Omega}_{ij}|^q \leq s_0 \right\}, \tag{2.3}$$

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

Under the situation when the observations are i.i.d. sub-Gaussian and the underlying $\boldsymbol{\Omega}$ belongs to $\mathcal{U}(q, s_0(p), M(p))$, Cai et al. (2011) establish consistency of $\widehat{\boldsymbol{\Omega}}_{\text{CLIME}}$ when q, $s_0(p)$ and $M(p)$ satisfy $M^{2-2q} s_0 (\log p/n)^{(1-q)/2} \to 0$; see Theorem 1(a) therein.

In the high-frequency setting, due to stochastic volatility, leverage effect etc., the returns are *not* i.i.d., so the results in Cai et al. (2011) do not apply. Before we discuss how to tackle this difficulty, we remark that the sparsity assumption $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} \in \mathcal{U}(q, s_0, M)$ appears to be reasonable in financial applications. For example, if the returns are assumed to

[1] This is related to the phenomenon that the volatility process is often found to be nearly unit root, in which case the one-step ahead prediction is approximately the current value. The assumption is also used in Fan et al. (2012a) and Dai et al. (2019), among others.

follow a (conditional) multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$, then the (i,j)th element in $\boldsymbol{\Omega}$ being 0 is equivalent to that the returns of the ith and jth assets are conditionally independent given the other asset returns. For stocks in different sectors, many pairs might be conditionally independent or only weakly dependent.

Now we discuss how to adopt CLIME to the high-frequency setting. Our goal is to estimate $\boldsymbol{\Sigma}_{\text{ICV}}$, or, more precisely, its inverse $\boldsymbol{\Omega}_{\text{ICV}} \coloneqq \boldsymbol{\Sigma}_{\text{ICV}}^{-1}$. In order to apply CLIME, we need to decide which $\widehat{\boldsymbol{\Sigma}}$ to use in (2.2).

When the true log-prices are observed, one of the most commonly used estimators for $\boldsymbol{\Sigma}_{\text{ICV}}$ is the realized covariance (RCV) matrix. Specifically, for each asset i, suppose the observations at stage n are $\left(X^i_{t^{i,n}_\ell}\right)$, 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

$$t^{i,n}_\ell \equiv t^n_\ell \quad \text{for all } i = 1, \ldots, p, \tag{2.4}$$

which, in the simplest equidistant setting, reduces to

$$t^{i,n}_\ell = t^n_\ell = \ell/n, \quad \ell = 0, 1, \ldots, n. \tag{2.5}$$

In the synchronous observation case (2.4), for $\ell = 1, \ldots, n$, let

$$\Delta \mathbf{X}_\ell \coloneqq \mathbf{X}_{t^n_\ell} - \mathbf{X}_{t^n_{\ell-1}}$$

be the log-return vector over the time interval $[t^n_{\ell-1}, t^n_\ell]$. The RCV matrix is defined as

$$\widehat{\boldsymbol{\Sigma}}_{\text{RCV}} = \sum_{\ell=1}^{n} \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^\intercal. \tag{2.6}$$

We are now ready to define the *constrained l_1-minimization for inverse matrix estimation with stochastic volatility* (CLIME-SV), $\widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SV}}$:

$$\widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SV}} \coloneqq \arg\min_{\boldsymbol{\Omega}'} \|\boldsymbol{\Omega}'\|_1 \text{ subject to } \|\widehat{\boldsymbol{\Sigma}}_{\text{RCV}} \boldsymbol{\Omega}' - \mathbf{I}\|_\infty \le \lambda. \tag{2.7}$$

The resulting MVP estimator is given by

$$\widehat{\boldsymbol{w}}_{\text{CLIME}-\text{SV}} = \frac{\widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SV}} \mathbf{1}}{\mathbf{1}^\intercal \widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SV}} \mathbf{1}}, \tag{2.8}$$

which is associated with a risk of

$$R_{\text{CLIME}-\text{SV}} = \widehat{\boldsymbol{w}}^\intercal_{\text{CLIME}-\text{SV}} \boldsymbol{\Sigma}_{\text{ICV}} \widehat{\boldsymbol{w}}_{\text{CLIME}-\text{SV}} = \frac{(\widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SV}} \mathbf{1})^\intercal \boldsymbol{\Sigma}_{\text{ICV}} (\widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SV}} \mathbf{1})}{(\mathbf{1}^\intercal \widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SV}} \mathbf{1})^2}. \tag{2.9}$$

Next we discuss the theoretical properties of this portfolio estimator. We make the following assumptions.

Assumption A.

(A.i) There exists $\delta \in (0, 1)$ such that for all p, $\delta \le \lambda_{\min}(\boldsymbol{\Sigma}_{\text{ICV}}) \le \lambda_{\max}(\boldsymbol{\Sigma}_{\text{ICV}}) < 1/\delta$ almost surely, where for any symmetric matrix A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ stand for its smallest and largest eigenvalues, respectively.

(A.ii) The drift process is such that $|\mu^i_t| \le C_\mu$ for some constant $C_\mu < \infty$ for all $i = 1, \ldots, p$ and $t \in [0, 1]$ almost surely.

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

(A.iv) The observation times t^n_ℓ under the synchronous setting (2.4) satisfy that there exists constant C_Δ such that

$$\sup_n \max_{1 \le \ell \le n} n|t^n_\ell - t^n_{\ell-1}| \le C_\Delta < \infty. \tag{2.10}$$

(A.v) p and n satisfy that $\log p/n \to 0$.

Remark 1. Assumptions (A.ii) and (A.iii) are standard in the high frequency literature, one assuming bounded drift, the other assuming bounded volatility. One may relax the bounded volatility assumption by using techniques similar to Lemma 3 of Fan et al. (2012a). Assumption (A.iv) is also widely adopted, although one may need to apply synchronization techniques such as the previous tick (Zhang, 2011) and refresh time (Barndorff-Nielsen et al., 2011) in order to synchronize the observations. Assumption (A.v) implies that the number of assets, p, can grow faster than the observation frequency.

Theorem 1. *Suppose that $(\mathbf{X}_t)$ satisfies (2.1) and the underlying precision matrix $\boldsymbol{\Omega}_{\text{ICV}} = \boldsymbol{\Sigma}^{-1}_{\text{ICV}} \in \mathcal{U}(q, s_0, M)$. Under Assumptions (A.i)–(A.v), with $\lambda = \eta M \sqrt{\log p/n}$ for some $\eta > 0$, there exist constants $C_1, C_2 > 0$ such that*

$$P\left(\frac{R_{\text{CLIME}-\text{SV}}}{R_{\min}} - 1 = O\left(M^{2-2q} s_0 \left((\log p/n)^{(1-q)/2}\right)\right)\right) \ge 1 - \frac{C_1}{p^{C_2\eta^2-2}},$$

where $R_{\text{CLIME}-\text{SV}}$ is defined in (2.9) and $R_{\min} = 1/(\mathbf{1}^\intercal \boldsymbol{\Omega}_{\text{ICV}} \mathbf{1})$.

Remark 2. 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 3. The case where observations are i.i.d. is 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.

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 $\left(X^i_{t^{i,n}_\ell}\right)$, for each asset i, the observations at stage n are

$$Y^i_{t^{i,n}_\ell} = X^i_{t^{i,n}_\ell} + \varepsilon^i_\ell, \tag{2.11}$$

where ε^i_ℓ's represent microstructure noise. In this case, if one simply plugs $\left(Y^i_{t^{i,n}_\ell}\right)$ into the formula of RCV in (2.6), 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), Podolskij and Vetter (2009) and Jacod et al. (2019)), realized kernels (RK, Barndorff-Nielsen et al. (2008)), quasi-maximum likelihood estimator (QMLE, Xiu (2010)), estimated-price realized volatility (ERV, Li et al. (2016)) and the unified volatility estimator (UV, Li et al. (2018)). 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 (2.5) (such as data sampled every minute). Asynchronicity can be dealt with by using existing data synchronization techniques. Compared with microstructure noise, asynchronicity is less an issue; see, for example, Section 2.4 in Xia and Zheng (2018).

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

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

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

$$\widehat{\mathbf{\Sigma}}_{\mathrm{PAV}} = \frac{12}{\theta\sqrt{n}}\sum_{k=0}^{n-k_n+1}\overline{\mathbf{Y}}^n_k\cdot(\overline{\mathbf{Y}}^n_k)^\intercal - \frac{6}{\theta^2 n}\operatorname{diag}\left(\sum_{k=1}^{n}(\Delta Y^i_{t_k})^2\right)_{i=1,\ldots,p}. \tag{2.12}$$

We now define the *constrained l_1-minimization for inverse matrix estimation with stochastic volatility and microstructure noise* (CLIME-SVMN), $\widehat{\mathbf{\Omega}}_{\mathrm{CLIME-SVMN}}$, as

$$\widehat{\mathbf{\Omega}}_{\mathrm{CLIME-SVMN}} := \arg\min_{\mathbf{\Omega}'}\|\mathbf{\Omega}'\|_1 \text{ subject to } \|\widehat{\mathbf{\Sigma}}_{\mathrm{PAV}}\mathbf{\Omega}' - \mathbf{I}\|_\infty \le \lambda. \tag{2.13}$$

Correspondingly, we define the estimated MVP, $\widehat{\boldsymbol{w}}_{\mathrm{CLIME-SVMN}}$, by replacing $\widehat{\mathbf{\Omega}}_{\mathrm{CLIME-SV}}$ in (2.8) with $\widehat{\mathbf{\Omega}}_{\mathrm{CLIME-SVMN}}$. Its associated risk, $R_{\mathrm{CLIME-SVMN}}$, is given by replacing $\widehat{\boldsymbol{w}}_{\mathrm{CLIME-SV}}$ and $\widehat{\mathbf{\Omega}}_{\mathrm{CLIME-SV}}$ in (2.9) with $\widehat{\boldsymbol{w}}_{\mathrm{CLIME-SVMN}}$ and $\widehat{\mathbf{\Omega}}_{\mathrm{CLIME-SVMN}}$, respectively.

Now we state the assumptions that we need in order to establish the statistical properties of the portfolio $\widehat{\boldsymbol{w}}_{\mathrm{CLIME-SVMN}}$.

Assumption B.

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

(B.ii) For any $\gamma \in \mathbb{R}$,

$$E(\exp(\gamma\varepsilon^i_\ell)) \le \exp(C\gamma^2),$$

where C is a fixed constant. Suppose also that there exists $C_\varepsilon > 0$ such that $\mathrm{Var}(\varepsilon^i_\ell) \le C_\varepsilon$ for all i and ℓ.

(B.iii) p and n satisfy that $\log p/\sqrt{n} \to 0$.

Theorem 2. *Suppose that $(\mathbf{X}_t)$ satisfies (2.1) and $\mathbf{\Omega}_{\mathrm{ICV}} = \mathbf{\Sigma}^{-1}_{\mathrm{ICV}} \in \mathcal{U}(q, s_0, M)$. Under Assumptions (A.i)–(A.iv) and (B.i)–(B.iii), with $\lambda = \eta M\sqrt{\log p}/n^{1/4}$ for some $\eta > 0$, there exist constants $C_3, C_4 > 0$ such that*

$$P\left(\frac{R_{\mathrm{CLIME-SVMN}}}{R_{\min}} - 1 = O\left(M^{2-2q}s_0\left((\log p)^{(1-q)/2}/n^{(1-q)/4}\right)\right)\right) \ge 1 - \frac{C_3}{p^{C_4\eta^2-2}}.$$

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.

2.4. Estimating the minimum risk

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

2.4.1. Assuming sparsity: using CLIME based estimator

The CLIME based estimators we considered in Sections 2.2 and 2.3 also lead to estimators of the minimum risk $R_{\min}$. Recall that $R_{\min} = 1/\mathbf{1}^{\intercal}\boldsymbol{\Omega}_{\text{ICV}}\mathbf{1}$. Our estimators are, when working under high-frequency setting without noise,

$$\widehat{R}_{\text{CLIME}-\text{SV}} = \frac{1}{\mathbf{1}^{\intercal}\widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SV}}\mathbf{1}}, \tag{2.14}$$

where, recall that, $\widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SV}}$ is defined in (2.7). Similarly, if microstructure noise is present, then the minimum risk estimator is defined as

$$\widehat{R}_{\text{CLIME}-\text{SVMN}} = \frac{1}{\mathbf{1}^{\intercal}\widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SVMN}}\mathbf{1}}, \tag{2.15}$$

where $\widehat{\boldsymbol{\Omega}}_{\text{CLIME}-\text{SVMN}}$ is defined in (2.13).

We have the following results for these estimators.

Theorem 3.

(i) Under the assumptions in Theorem 1, the minimum risk estimator $\widehat{R}_{\text{CLIME}-\text{SV}}$ defined in (2.14) satisfies that

$$P\left(\frac{\widehat{R}_{\text{CLIME}-\text{SV}}}{R_{\min}} - 1 = O\left(M^{2-2q}s_0\left((\log p/n)^{(1-q)/2}\right)\right)\right) \geq 1 - \frac{C_1}{p^{C_2\eta^2-2}}.$$

(ii) Under the assumptions in Theorem 2, the minimum risk estimator $\widehat{R}_{\text{CLIME}-\text{SVMN}}$ defined in (2.15) satisfies that

$$P\left(\frac{\widehat{R}_{\text{CLIME}-\text{SVMN}}}{R_{\min}} - 1 = O\left(M^{2-2q}s_0\left((\log p)^{(1-q)/2}/n^{(1-q)/4}\right)\right)\right) \geq 1 - \frac{C_3}{p^{C_4\eta^2-2}}.$$

2.4.2. Without sparsity assumption: low-frequency i.i.d. returns

CLIME based estimators work under mild sparsity assumptions on the precision matrix. Now we propose another estimator of the minimum risk, which does not rely on such sparsity assumptions, 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 over a long time period.

More specifically, suppose that we observe n i.i.d. returns $\mathbf{X}_1, \ldots, \mathbf{X}_n$ (possibly at 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. The corresponding *perceived* risk is $\widehat{R}_p = \boldsymbol{w}_p^{\intercal}\mathbf{S}\boldsymbol{w}_p$. We have the following result on the relationship between $\widehat{R}_p$ and the minimum risk $R_{\min}$, based on which a consistent estimator of the minimum risk is constructed.

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

$$\left|\frac{\widehat{R}_p}{R_{\min}} - (1 - \rho_n)\right| \xrightarrow{p} 0. \tag{2.16}$$

Therefore, if we define

$$\widehat{R}_{\min} = \frac{1}{1 - \rho_n}\widehat{R}_p, \tag{2.17}$$

then

$$\frac{\widehat{R}_{\min}}{R_{\min}} \xrightarrow{p} 1. \tag{2.18}$$

Furthermore, we have

$$\sqrt{n-p}\left(\frac{\widehat{R}_{\min}}{R_{\min}}-1\right)\Rightarrow N(0,2). \tag{2.19}$$

The convergence (2.19) actually shows "blessing" of dimensionality: the higher the dimension, the more accurate the estimation.

The convergence (2.16) explains why in Fig. 1 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 and 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

$$\frac{R(\boldsymbol{w}_p)}{R_{\min}}\xrightarrow{p}\frac{1}{1-\rho}, \tag{2.20}$$

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

3. Estimation with factors

In practice, stock returns often exhibit a factor structure. Prominent examples include CAPM (Sharpe, 1964), Fama–French Three Factor and Five Factor models (Fama and French, 1992, 2015). Recently, there have also been studies on factor modeling for high-frequency returns; see Aït-Sahalia and Xiu (2017), Dai et al. (2019), Pelger (2019). In this section, we extend our approach to accommodate such a structure.

Let $\mathbf{r}_k=(r_{k,1},\ldots,r_{k,p})^{\intercal}$, $k=1,\ldots,n$, be asset returns, which can be either high-frequency such as 5-minute returns or low-frequency like daily or weekly returns. We assume that $(\mathbf{r}_k)$ admits a factor structure as follows:

$$\mathbf{r}_k=\alpha+\Gamma\mathbf{f}_k+\mathbf{z}_k, \tag{3.1}$$

where α is a $p\times 1$ unknown vector, Γ is a $p\times m$ unknown coefficient matrix, $\mathbf{f}_k=(f_{k,1},\ldots,f_{k,m})^{\intercal}$ are factor returns, and $\mathbf{z}_k$ is $p\times 1$ random vector with mean 0 and covariance matrix $\boldsymbol{\Sigma}_{k,0}=(\sigma_{ij}^{k,0})$. We assume that for each k, $\mathbf{f}_k$'s and $\mathbf{z}_k$'s are independent, and the pairs $(\mathbf{r}_k,\mathbf{f}_k)$, $k=1,\ldots,n$, are mutually independent. Let $\boldsymbol{\Sigma}_{\mathbf{r},k}=\mathrm{Cov}(\mathbf{r}_k)$ and $\boldsymbol{\Sigma}_{\mathbf{f},k}=\mathrm{Cov}(\mathbf{f}_k)$. Note that we allow both $\boldsymbol{\Sigma}_{\mathbf{r},k}$ and $\boldsymbol{\Sigma}_{\mathbf{f},k}$ to depend on the time index k, to accommodate stochastic (co-)volatility. Because of such a feature, it is impossible to estimate individual $\boldsymbol{\Sigma}_{\mathbf{r},k}$ or $\boldsymbol{\Sigma}_{\mathbf{f},k}$, but it is possible to estimate their means, namely, $\boldsymbol{\Sigma}_{\mathbf{r}}=\frac{1}{n}\sum_{k=1}^n\boldsymbol{\Sigma}_{\mathbf{r},k}$, $\boldsymbol{\Sigma}_{\mathbf{f}}=\frac{1}{n}\sum_{k=1}^n\boldsymbol{\Sigma}_{\mathbf{f},k}$, and $\boldsymbol{\Sigma}_0=\frac{1}{n}\sum_{k=1}^n\boldsymbol{\Sigma}_{k,0}$ and the corresponding $\boldsymbol{\Omega}_0=\boldsymbol{\Sigma}_0^{-1}$. The estimation procedure requires consistently estimating the means of $(\mathbf{r}_k,\mathbf{f}_k)$, which entails the condition that $E(\mathbf{f}_k)$'s are the same for all k.

To estimate the model, we first calculate the least squares estimators of α and Γ, denoted by $\hat{\alpha}$ and $\widehat{\Gamma}$ respectively. The residuals are $\mathbf{e}_k:=\mathbf{r}_k-\hat{\alpha}-\widehat{\Gamma}\mathbf{f}_k$. We then obtain a CLIME estimator of $\boldsymbol{\Omega}_0$, denoted by $\widehat{\boldsymbol{\Omega}}_0$, based on the residuals. Specifically,

$$\widehat{\boldsymbol{\Omega}}_0:=\arg\min_{\boldsymbol{\Omega}'}\|\boldsymbol{\Omega}'\|_1\ \text{subject to}\ \|\mathbf{S}_{\mathbf{e}}\boldsymbol{\Omega}'-\mathbf{I}\|_\infty\le\lambda, \tag{3.2}$$

where $\mathbf{S}_{\mathbf{e}}$ is the sample covariance matrix of residuals $(\mathbf{e}_1,\ldots,\mathbf{e}_n)$.

Our goal is to estimate the precision matrix $\boldsymbol{\Omega}_{\mathbf{r}}:=\boldsymbol{\Sigma}_{\mathbf{r}}^{-1}$. To do so, note that $\boldsymbol{\Sigma}_{\mathbf{r}}=\Gamma\boldsymbol{\Sigma}_{\mathbf{f}}\Gamma^{\intercal}+\boldsymbol{\Sigma}_0$, it follows that

$$\boldsymbol{\Omega}_{\mathbf{r}}=(\Gamma\boldsymbol{\Sigma}_{\mathbf{f}}\Gamma^{\intercal}+\boldsymbol{\Sigma}_0)^{-1}=\boldsymbol{\Omega}_0-\boldsymbol{\Omega}_0\Gamma(\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}+\Gamma^{\intercal}\boldsymbol{\Omega}_0\Gamma)^{-1}\Gamma^{\intercal}\boldsymbol{\Omega}_0. \tag{3.3}$$

Therefore, we can define the *constrained l_1-minimization for inverse matrix estimation adjusted for factor* (CLIME-F), $\widehat{\boldsymbol{\Omega}}_{\mathrm{CLIME-F}}$, as

$$\widehat{\boldsymbol{\Omega}}_{\mathrm{CLIME-F}}=\widehat{\boldsymbol{\Omega}}_0-\widehat{\boldsymbol{\Omega}}_0\widehat{\Gamma}(\mathbf{S}_{\mathbf{f}}^{-1}+\widehat{\Gamma}^{\intercal}\widehat{\boldsymbol{\Omega}}_0\widehat{\Gamma})^{-1}\widehat{\Gamma}^{\intercal}\widehat{\boldsymbol{\Omega}}_0, \tag{3.4}$$

where $\mathbf{S}_{\mathbf{f}}$ is the sample covariance matrix of $(\mathbf{f}_1,\ldots,\mathbf{f}_n)$. From here the MVP estimator is

$$\widehat{\boldsymbol{w}}_{\mathrm{CLIME-F}}=\frac{\widehat{\boldsymbol{\Omega}}_{\mathrm{CLIME-F}}\mathbf{1}}{\mathbf{1}^{\intercal}\widehat{\boldsymbol{\Omega}}_{\mathrm{CLIME-F}}\mathbf{1}}. \tag{3.5}$$

To establish the statistical properties for this estimator, the following assumptions are made.

Assumption C.

(C.i) Let the number of factors, m, be fixed, and $\log(p)=o(n)$. Moreover, there exist $\eta>0$ and $K>0$ such that for all $i=1,\ldots,m$, $j=1,\ldots,p$, and $k=1,\ldots,n$,

$$E(\exp(\eta f_{ki}^2))\le K,\quad E(\exp(\eta z_{kj}^2/\sigma_{jj}^{k,0}))\le K,\quad\text{and}\quad\sigma_{jj}^{k,0}\le K.$$

(C.ii) $\boldsymbol{\Sigma}_{\mathbf{f}} = \frac{1}{n}\sum_{k=1}^{n} \mathrm{Cov}(\mathbf{f}_k)$ is of full rank, and there exists $K_{\mathbf{f}} > 0$ such that $\|\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\|_{L_\infty} \le K_{\mathbf{f}}$, where for any matrix $A = (a_{ij})$, $\|A\|_{L_\infty} := \max_i \sum_j |a_{ij}|$.

(C.iii) There exists $\delta \in (0, 1)$ such that for all p, $\delta \le \lambda_{\min}(\boldsymbol{\Sigma}_0) \le \lambda_{\max}(\boldsymbol{\Sigma}_0) < 1/\delta$.

(C.iv) For all p, the eigenvalues of $p^{-1}\Gamma^\intercal\Gamma$ are bounded away from 0 and ∞.

Theorem 4. *Suppose that the returns $(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ follow (3.1) and the precision matrix $\boldsymbol{\Omega}_0 \in \mathcal{U}(q, s_0, M)$. Under Assumptions (C.i)–(C.iv), with probability greater than $1 - O(p^{-1})$, we have, for some constant $C_5 > 0$,*

$$\|\widehat{\Gamma} - \Gamma\|_\infty \le C_5 \left(\frac{\log p}{n}\right)^{1/2}. \tag{3.6}$$

Moreover, with $\lambda = C_6 M\sqrt{\log p/n}$ for some sufficiently large positive constant C_6, with probability greater than $1 - O(p^{-1})$, we have, for some constants $C_7, C_8 > 0$,

$$\|\widehat{\boldsymbol{\Omega}}_0 - \boldsymbol{\Omega}_0\|_2 \le C_7 M^{1-q} s_0 \left(\frac{\log p}{n}\right)^{(1-q)/2}, \tag{3.7}$$

and

$$\|\widehat{\boldsymbol{\Omega}}_{\mathrm{CLIME-F}} - \boldsymbol{\Omega}_{\mathbf{r}}\|_2 \le C_8 M^{1-q} s_0 \left(\frac{\log p}{n}\right)^{(1-q)/2}, \tag{3.8}$$

where for any symmetric matrix A, $\|A\|_2$ stands for its spectral norm.

Remark 6. When $(\mathbf{r}_k, \mathbf{f}_k)$, $k = 1, \ldots, n$, are i.i.d., we have $\boldsymbol{\Sigma}_{\mathbf{r},k} \equiv \boldsymbol{\Sigma}_{\mathbf{r}}$ and $\boldsymbol{\Sigma}_{\mathbf{f},k} \equiv \boldsymbol{\Sigma}_{\mathbf{f}}$, and Theorem 4 readily applies. We do not impose such a strong assumption and allow both $\boldsymbol{\Sigma}_{\mathbf{r},k}$ and $\boldsymbol{\Sigma}_{\mathbf{f},k}$ to change over k, to accommodate the stochastic (co-)volatility in high-frequency returns.

4. Simulation studies

We simulate the p-dimensional log-price process $(\mathbf{X}_t)$ from a stochastic factor model:

$$d\mathbf{X}_t = \Gamma\big(\boldsymbol{\mu}_{\mathbf{f}}\, dt + \gamma_t \boldsymbol{\Sigma}_{\mathbf{f}}^{1/2}\, d\boldsymbol{W}_t^{(1)}\big) + \gamma_t \boldsymbol{\Sigma}_0^{1/2}\, d\boldsymbol{W}_t^{(2)}, \quad \text{for } t \in [0, 1], \tag{4.1}$$

where $\boldsymbol{\mu}_{\mathbf{f}}$ and $\boldsymbol{\Sigma}_{\mathbf{f}}$ are the mean vector and covariance matrix of an m-dimensional factor return, $(\boldsymbol{W}_t^{(1)})$ and $(\boldsymbol{W}_t^{(2)})$ are independent m-dimensional and p-dimensional standard Brownian motions, and (γ_t) is a stochastic U-shaped process that obeys

$$d\gamma_t = -\rho(\gamma_t - \varphi_t)\, dt + \sigma\, dB_t,$$

where $\rho = 10$, $\sigma = 0.05$,

$$\varphi_t = \sqrt{1 + 0.8\cos(2\pi t)}, \quad \text{for } t \in [0, 1],$$

and (B_t) is a standard Brownian motion that is independent of $(\boldsymbol{W}_t^{(1)}, \boldsymbol{W}_t^{(2)})$. Under such a setting, the ICV matrix is

$$\boldsymbol{\Sigma}_{\mathrm{ICV}} = \left(\int_0^1 \gamma_t^2 dt\right)\boldsymbol{\Sigma}, \text{ where } \boldsymbol{\Sigma} = \Gamma\boldsymbol{\Sigma}_{\mathbf{f}}\Gamma^\intercal + \boldsymbol{\Sigma}_0.$$

We now discuss the specifications of the parameters in (4.1). The parameter values are learned from real data as follows. We work with the daily returns in Year 2012 of randomly selected eighty S&P100 Index constituents that are used in our empirical studies (Section 5). The factors are taken to mimic the Fama–French three factors, and we set $\boldsymbol{\mu}_{\mathbf{f}}$ and $\boldsymbol{\Sigma}_{\mathbf{f}}$ to be the sample mean and sample covariance matrix of the daily factor returns in Year 2012. As to the coefficients Γ, they are taken to be the estimated coefficients by fitting the Fama–French three-factor model to the daily returns of the eighty stocks. Finally, we construct a CLIME estimator $\widehat{\boldsymbol{\Omega}}_0$ based on the residuals as in (3.2), and set $\boldsymbol{\Sigma}_0 = \widehat{\boldsymbol{\Omega}}_0^{-1}$.

We shall compare our proposed portfolios with several existing methods that are based on low-frequency data. We simulate 11 years' data, each year consisting of $T = 252$ days and each day $n = 391$ log-prices. The first year is taken as the initial training period and the remaining 10 years as the testing period. The log-prices are contaminated with simulated microstructure noise. Specifically, for each stock i, the noise is generated from $N(0, \sigma_{e,i}^2)$, where $\sigma_{e,i} \sim \mathrm{Unif}(0.0001, 0.0005)$.

We include the following portfolios in our comparison. All portfolios are constructed using a rolling window scheme, with different window lengths according to whether a portfolio is a low-frequency strategy or a high-frequency one. For methods using low-frequency data (methods (ii)–(vi)), daily returns during the immediate past 252 days are used to construct the covariance matrix estimator $\widehat{\boldsymbol{\Sigma}}$ or precision matrix estimator $\widehat{\boldsymbol{\Omega}}$, which is then plugged into (1.2), and the weights are re-calculated every 21 days. For methods using high-frequency data (methods (vii)–(ix)), the portfolio weights are estimated with returns from the immediate past 10 days and are re-estimated every 5 days.

Table 1
Simulation comparison based on 10 years testing period. For each portfolio, its annualized standard deviation (Ann. SD) is reported.

	Low Frequency			High Frequency	
Non-factor-based	EW	NLS	CLIME	CLIME-SV(5min)	CLIME-SVMN(1min)
Ann. SD	9.64%	6.29%	6.27%	6.08%	6.01%
Factor-based	FFL	NLSF	CLIME-F	CLIME-F(5min)	Oracle
Ann. SD	8.00%	6.18%	6.10%	5.96%	5.49%

Table 2
Annualized unconditional minimum risk estimates.

Estimator	$\widehat{R}_{\min}$ based on (2.17)	$\widehat{R}_{\text{CLIME}-\text{SVMN}}$	Oracle
Estimate (annualized)	5.51%	5.29%	5.49%

(i) Equally-weighted (EW) portfolio. This serves as an important benchmark portfolio as shown in DeMiguel et al. (2009b);
(ii) FFL portfolio. This is based on Fan et al. (2008) and utilizes the three factors from the data-generating process;
(iii) Nonlinear shrinkage (NLS) portfolio. This is based on Ledoit and Wolf (2012);
(iv) NLSF portfolio. This is based on Ledoit and Wolf (2017);
(v) CLIME portfolio. To construct this portfolio, daily returns are treated as i.i.d. observations and the original CLIME method is directly applied to estimate the precision matrix and the portfolio weight;
(vi) CLIME-F portfolio. This is based on Theorem 4 using daily returns and utilizing the three factors from the data-generating process;
(vii) CLIME-SV(5min) portfolio. This is based on Theorem 1 using 5-min returns;
(viii) CLIME-F(5min) portfolio: This is a high-frequency version of CLIME-F where the factor model is estimated using 5-min intra-day returns;
(ix) CLIME-SVMN(1min) portfolio. This is based on Theorem 2 using 1-min returns;
(x) 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.

Table 1 reports the annualized standard deviations of these portfolios based on 10 years of testing.
We observe from Table 1 that

- Among low-frequency methods, when the factor structure is not taken into consideration, CLIME performs substantially better than EW and is comparable to NLS;
- Our proposed high-frequency methods, CLIME-SV and CLIME-SVMN, work even better. In particular, CLIME-SVMN can be used on noisy high-frequency data and attains a substantially lower risk;
- Further incorporating factor structure helps. CLIME-F(5min) attains the lowest risk among all feasible portfolios.

Next, we examine the estimation of minimum risk. As we discussed in Section 2.1, the daily minimum risk is a random variable that changes from one day to another. Hence, we obtain one estimated $\widehat{R}_{\text{CLIME}-\text{SVMN}}$ for each day, and totally $252 \times 10 = 2520$ estimates. By the law of total variance and under the stationarity assumption, we can then calculate the unconditional daily minimum risk by taking the average of these estimates (ignoring the variation in the expected daily returns). Alternatively, we can use daily returns in the whole sample to estimate the (long-term) minimum risk based on Proposition 1 ($\widehat{R}_{\min}$ based on (2.17)). The results are summarized in Table 2.
We see from Table 2 that both estimators give quite accurate estimates.

5. Empirical studies

In this section, we conduct empirical studies and compare minimum variance portfolios constructed via different methods as well as the equally-weighted portfolio.

We work with S&P 100 Index constitutes during Years 2003–2013. For each year to be tested, we consider those stocks that are included in the Index during that year as well as the previous year. For example, when constructing portfolios for Year 2004, the stocks considered are constitutes that are common in Years 2003 and 2004, and for Year 2005, the stocks used would be those common constitutes for Years 2004 and 2005.[2] The observations are 1-min intra-day prices. For methods based on factor models, Fama–French three factors are the factors that are utilized.[3] In addition to the portfolios

[2] For stock pools selection, this is a bit forward looking. This helps to maintain the quality of data in test. Similar arrangements were used in other research works. Note that other than this, we use no future information. Our testing is fully out-of-sample.

[3] We thank the authors of Dai et al. (2019) for sharing their high-frequency (5-min) factors data with us.

Table 3
Portfolio comparison based on S&P100 Index constituents during Years 2003–2013. For each portfolio, its annualized standard deviation (Ann. SD) is reported.

	Low Frequency			High Frequency	
Non-factor-based	EW	NLS	CLIME	CLIME-SV(5min)	CLIME-SVMN(1min)
Ann. SD	14.84%	9.64%	9.52%	9.56%	9.25%
Factor-based	FFL	NLSF	CLIME-F	DLX(5min)	CLIME-F(5min)
Ann. SD	9.73%	9.50%	9.20%	9.52%	9.13%

Table 4
Comparison of (annualized) average monthly volatilities based on 15-min returns.

	Low Frequency			High Frequency	
Non-factor-based	EW	NLS	CLIME	CLIME-SV(5min)	CLIME-SVMN(1min)
Ann. Vol.	14.46%	11.04%	10.86%	10.19%	10.12%
Factor-based	FFL	NLSF	CLIME-F	DLX(5min)	CLIME-F(5min)
Ann. Vol.	10.84%	10.82%	10.60%	10.34%	9.92%

in the Simulation studies (except the Oracle portfolio which is infeasible in practice), we add the following portfolio, which represents the latest development in the literature:

- DLX(5min) portfolio: This is based on Dai et al. (2019) and uses the Fama–French three-factor model. 5-min intra-day observations from the immediate past 10 days with GICS codes at the industry group level are used to construct the covariance matrix estimator $\widehat{\Sigma}$, which is plugged into (1.2) to obtain the portfolio weights. The weights are re-estimated every 5 days.

Note that because this method depends on GICS codes, we cannot implement it in simulations.

Table 3 reports the annualized standard deviations of all the portfolios under comparison based on daily returns in the 10-year testing period.

From Table 3 we observe similar comparisons to Table 1. Specifically,

- Among low-frequency methods, when the factor structure is not taken into consideration, CLIME outperforms EW and NLS;
- With high-frequency data, CLIME-SVMN improves and attains a substantially lower risk;
- CLIME-F(5min) performs even better and attains the lowest risk.

As a reference, the minimum risk estimated based on (2.17) with daily returns is 8.28%. This indicates that our methods approach the minimum risk reasonably well.

Dai et al. (2019) use a slightly different measure for comparison, which is based on monthly realized volatilities. Specifically, for each portfolio, its monthly realized volatilities (based on 15-minute returns) are calculated and the average of monthly volatilities (after annualization) is reported. The comparison under this alternative measure is reported in Table 4.

We observe that our methods outperform the approach of Dai et al. (2019) based on such an alternative measure as well.

To practically implement MVP, an important issue is the possible presence of extreme positions, which sometimes calls for additional short-sale or gross-exposure constraints. For this reason, we analyze the maximum (absolute) weights of each portfolio during the testing period. Table 5 reports the medians and means of extreme positions for different methods.

We observe from Table 5 that our proposed methods appear to be advantageous in this regard. In both non-factor-based and factor-based panels, CLIME-based methods have milder extreme positions. An interesting observation is that factor-based methods lead to more severe extreme positions.

6. Conclusion

We propose estimators of the minimum variance portfolio in the high-dimensional setting based on high-frequency data. The desired risk convergence (1.8) for the estimated portfolios is obtained under certain sparsity assumptions on the precision matrix. We further propose consistent estimators of the minimum risk, one of which does not require sparsity.

Numerical studies demonstrate that our methods perform favorably. An important conclusion is that high-frequency data can add value to portfolio allocation.

Acknowledgments

We thank the Editor and two anonymous referees for their constructive suggestions. We also thank the participants of SoFiE 2016 and FERM 2016 conferences for helpful discussions on the preliminary versions of this paper. We are grateful

Table 5
Median and mean of max (absolute) weight for each portfolio in the testing period. Portfolio EW has evenly distributed positions of 0.012 on each asset, and is not included in the table.

Non-factor-based	Median	Mean
NLS	0.113	0.119
CLIME	0.071	0.074
CLIME-SV(5min)	0.069	0.070
CLIME-SVMN(1min)	0.089	0.089
Factor-based	**Median**	**Mean**
FFL	0.128	0.133
NLSF	0.158	0.163
CLIME-F	0.126	0.126
DLX(5min)	0.163	0.174
CLIME-F(5min)	0.114	0.118

to insightful comments from Robert Engle, Raymond Kan, Olivier Ledoit and Dacheng Xiu. We thank the authors of Dai et al. (2019) for sharing their high-frequency factors data with us. Research was partially supported by National Science Foundation, United States of America Grant DMS-1712735 and NIH, United States of America grants R01-GM129781 and R01-GM123056; and the RGC, Hong Kong grants GRF 16305315, GRF 16304317, GRF 16502014, GRF 16518716 and GRF 16502118 of the HKSAR.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2019.04.039.

References

Aït-Sahalia, Y., Xiu, D., 2017. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. J. Econometrics 201 (2), 384–399.
Ao, M., Li, Y., Zheng, X., 2018. Approaching mean-variance efficiency for large portfolios. Rev. Financial Stud. (in press).
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., 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.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. J. Econometrics 162, 149–169.
Basak, G.K., Jagannathan, R., Ma, T., 2009. Jackknife estimator for tracking error variance of optimal portfolios. Manage. Sci. 55 (6), 990–1002.
Cai, T.T., Liu, W., Luo, X., 2011. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. J. Amer. Statist. Assoc. 106, 594–607.
Chan, L.K., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: forecasting covariances and choosing the risk model. Rev. Financ. Stud. 12 (5), 937–974.
Christoffersen, P., 2012. Elements of Financial Risk Management. Academic Press, Oxford.
Clarke, R.G., De Silva, H., Thorley, S., 2006. Minimum-variance portfolios in the US equity market. J. Portfolio Manag. 33 (1), 10–24.
Dai, C., Lu, K., Xiu, D., 2019. Knowing factors or factor loadings, or neither? evaluating estimators of large covariance matrices with noisy and asynchronous data. J. Econometrics 208 (1), 43–79.
DeMiguel, V., Garlappi, L., Nogales, F.J., Uppal, R., 2009a. A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Manage. Sci. 55 (5), 798–812.
DeMiguel, V., Garlappi, L., Uppal, R., 2009b. Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? Rev. Financ. Stud. 22 (5), 1915–1953.
Ding, Y., Li, Y., Zheng, X., 2018. Estimating high dimensional minimum variance portfolio under factor model. Working paper.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. J. Finance 47 (2), 427–465.
Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116 (1), 1–22.
Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. J. Econometrics 147 (1), 186–197.
Fan, J., Li, Y., Yu, K., 2012a. Vast volatility matrix estimation using high-frequency data for portfolio selection. J. Amer. Statist. Assoc. 107 (497), 412–428.
Fan, J., Zhang, J., Yu, K., 2012b. Vast portfolio selection with gross-exposure constraints. J. Amer. Statist. Assoc. 107 (498), 592–606.
Gloter, A., Jacod, J., 2001. Diffusions with measurement errors. I. local asymptotic normality. ESAIM Probab. Stat. 5, 225–242.
Haugen, R.A., Baker, N.L., 1991. The efficient market inefficiency of capitalization-weighted stock portfolios. J. Portfolio Manag. 17 (3), 35–40.
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Process. Appl. 119 (7), 2249–2276.
Jacod, J., Li, Y., Zheng, X., 2019. Estimating the integrated volatility with tick observations. J. Econometrics 208 (1), 80–100.
Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. J. Finance 58 (4), 1651–1684.
Kim, D., Wang, Y., Zou, J., 2016. Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. Stochastic Process. Appl. 126 (11), 3527–3577.
Ledoit, O., Wolf, M., 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. Ann. Statist. 40 (2), 1024–1060.
Ledoit, O., Wolf, M., 2017. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. Rev. Financ. Stud. 30 (12), 4349–4388.
Li, Y., Xie, S., Zheng, X., 2016. Efficient estimation of integrated volatility incorporating trading information. J. Econometrics 195 (1), 33–50.
Li, Y., Zhang, Z., Li, Y., 2018. A unified approach to volatility estimation in the presence of both rounding and random market microstructure noise. J. Econometrics 203 (2), 187–222.
Markowitz, H., 1952. Portfolio selection. J. Finance 7 (1), 77–91.

Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. J. Financ. Econ. 8 (4), 323–361.
Pelger, M., 2019. Large-dimensional factor modeling based on high-frequency observations. J. Econometrics 208 (1), 23–42.
Podolskij, M., Vetter, M., 2009. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli 15 (3), 634–658.
Schwartz, T., 2000. How to beat the S&P 500 with portfolio optimization. Working paper.
Sharpe, W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. J. Finance 19 (3), 425–442.
Tao, M., Wang, Y., Chen, X., 2013. Fast convergence rates in estimating large volatility matrices using high-frequency financial data. Econometric Theory 29 (4), 838–856.
Tao, M., Wang, Y., Yao, Q., Zou, J., 2011. Large volatility matrix inference via combining low-frequency and high-frequency approaches. J. Amer. Statist. Assoc. 106, 1025–1040.
Wang, Y., Zou, J., 2010. Vast volatility matrix estimation for high-frequency financial data. Ann. Statist. 38 (2), 943–978.
Xia, N., Zheng, X., 2018. On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations. Ann. Statist. 46 (2), 500–525.
Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159 (1), 235–250.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12 (6), 1019–1043.
Zhang, L., 2011. Estimating covariation: epps effect, microstructure noise. J. Econometrics 160 (1), 33–47.
Zhang, L., Mykland, P.A., 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.
Zheng, X., Li, Y., 2011. On the estimation of integrated covariance matrices of high dimensional diffusion processes. Ann. Statist. 39 (6), 3121–3151.