Dynamic Realized Minimum Variance Portfolio Models

Donggyu Kim
College of Business,
Korea Advanced Institute of Science and Technology (KAIST)
and
Minseog Oh
College of Business,
Korea Advanced Institute of Science and Technology (KAIST)

January 7, 2024

Abstract

This paper introduces a dynamic minimum variance portfolio (MVP) model using nonlinear volatility dynamic models, based on high-frequency financial data. Specifically, we impose an autoregressive dynamic structure on MVP processes, which helps capture the MVP dynamics directly. To evaluate the dynamic MVP model, we estimate the inverse volatility matrix using the constrained ℓ_1-minimization for inverse matrix estimation (CLIME) and calculate daily realized non-normalized MVP weights. Based on the realized non-normalized MVP weight estimator, we propose the dynamic MVP model, which we call the dynamic realized minimum variance portfolio (DR-MVP) model. To estimate a large number of parameters, we employ the least absolute shrinkage and selection operator (LASSO) and predict the future MVP and establish its asymptotic properties. Using high-frequency trading data, we apply the proposed method to MVP prediction.

Keywords: CLIME, high-frequency data, LASSO, nonlinear dynamics, inverse matrix

1 Introduction

The minimum variance portfolio (MVP) has received growing attention (DeMiguel et al., 2009; Fan et al., 2012; Ledoit and Wolf, 2017). Unlike the mean-variance portfolio proposed in Markowitz (1952), the MVP avoids the difficulty of estimating expected returns (Merton, 1980) and only requires accurate covariance matrix estimators. Furthermore, empirical studies have found that the MVP can enjoy both lower risk and higher return, compared with some benchmark portfolios (Chan et al., 1999; Clarke et al., 2006; Haugen and Baker, 1991; Jagannathan and Ma, 2003).

Several well-performing non-parametric realized volatility estimators have been developed since high-frequency financial data became available. Examples of the finite number of assets include two-time scale realized volatility (Zhang et al., 2005), multi-scale realized volatility (Zhang, 2006, 2011), wavelet estimator (Fan and Wang, 2007), kernel realized volatility (Barndorff-Nielsen et al., 2008, 2011), pre-averaging realized volatility (Christensen et al., 2010; Jacod et al., 2009), quasi-maximum likelihood estimator (Aït-Sahalia et al., 2010; Xiu, 2010), local method of moments (Bibinger et al., 2014), and robust pre-averaging realized volatility (Fan and Kim, 2018; Shin et al., 2023). On the other hand, estimation methods for large volatility matrices have been developed based on the sparse volatility matrix structure and approximate factor models (Aït-Sahalia and Xiu, 2017; Dai et al., 2019; Fan et al., 2012; Kim et al., 2018a,b; Pelger, 2019; Tao et al., 2013). Using these non-parametric volatility matrix estimation procedures, we can accurately estimate large volatility matrices within a relatively short time period, such as one day or a few days. This makes it possible to estimate past optimal MVP weights. For example, Fan et al. (2012) studied the MVP under the gross-exposure constraint, and Cai et al. (2020) proposed the estimation of high-dimensional MVP using the constrained ℓ_1-minimization for inverse matrix estimation (CLIME) (Cai et al., 2011).

In practice, we often need to predict future MVPs given current information. To predict the MVP, we often first predict the large volatility matrix and then we solve the minimum variance portfolio problem with the predicted large volatility matrix estimator (Kim and Fan, 2019; Shin et al., 2021). However, this approach does not reflect the MVP dynamics directly, which may cause some loss of explanatory power from them. To check the MVP dynamics, we draw the box-plots for the autocorrelations for each non-normalized MVP weight in Figure 1, where the non-parametric MVP weight estimation method is presented in Section 3.1. The horizontal red dotted line in Figure 1 indicates the significant autocorrelation level with 5% significance. From Figure 1, we find that the weight has autoregressive dynamics. Thus, by directly modeling these dynamics, we may be able to better explain the MVP dynamics. This fact leads us to develop an MVP dynamic model.
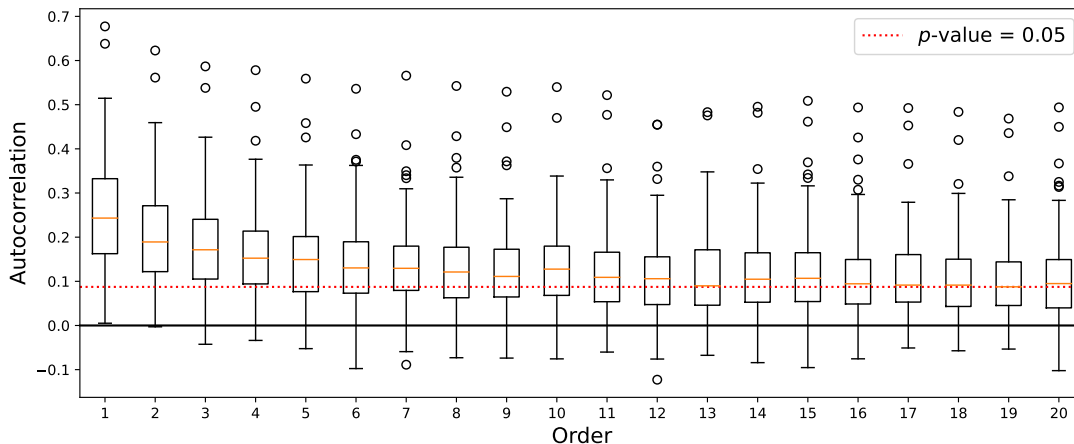

Figure 1: Box-plots of the first- to twentieth-order autocorrelations for each non-normalized MVP weight.

In this paper, we propose a dynamic MVP model based on high-frequency financial data. Specifically, we first investigate a non-parametric MVP estimation procedure. To tackle the curse of dimensionality, we impose a sparse structure on the inverse matrices of daily integrated volatility matrices. Using high-frequency financial data, we estimate daily integrated volatility matrices and then, to accommodate the sparsity, employ the CLIME procedure (Cai et al., 2011). With the inverse matrix estimator, we estimate non-

normalized MVP weights and establish their asymptotic properties. Then, to capture the autoregressive dynamic structure of these weights, we propose a dynamic MVP model. Since the proposed model explains low-frequency dynamics based on high-frequency financial data, we connect the low-frequency dynamics with a continuous diffusion process. That is, the proposed model has the rigorous mathematical background when it comes to incorporating high-frequency financial data to analyze low-frequency dynamics. We call this the dynamic realized minimum variance portfolio (DR-MVP) model. To evaluate the proposed DR-MVP model, we need to estimate a large number of parameters that come from a large number of assets and long period time series. To handle this curse of dimensionality, we assume that the model parameters are sparse and employ an ℓ_1-penalty, such as the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996). Based on the predicted non-normalized MVP weights, we estimate the conditional expected future MVP and show its consistency. In the empirical study, we find the benefit of adopting the DR-MVP model.

The rest of the paper is organized as follows. Section 2 introduces the DR-MVP model based on high-dimensional Itô diffusion processes and investigates its properties. Section 3 proposes the non-parametric MVP estimation procedure and the parameter estimation method for the high-dimensional DR-MVP model, and establishes its asymptotic properties. In Section 4, we conduct a simulation study to check the finite sample performance of the proposed estimator. In Section 5, we apply the estimation method to high-frequency trading data. The conclusion is presented in Section 6, and we collect the proofs in the online Appendix.

2 Realized Minimum Variance Portfolio

Let $\mathbf{X}_t = (X_{1,t}, \ldots, X_{p,t})^\top$ be the vector of true log-prices of p assets at time t. The log-price follows the following continuous diffusion process:

$$dX_t = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\mathcal{B}_t, \tag{2.1}$$

where $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ are drift and instantaneous volatility processes, and $\mathcal{B}_t$ is a p-dimensional standard Brownian motion. Stochastic processes $\mathbf{X}_t$, $\boldsymbol{\mu}_t$, and $\boldsymbol{\sigma}_t$ are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in [0, \infty)\}, P)$ with filtration $\mathcal{F}_t$ satisfying the usual conditions, such as càdlàg and adapted processes. We denote $\boldsymbol{\Sigma}_t = \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^\top$; that is, $\boldsymbol{\sigma}_t$ is square root of positive semidefinite spot covariance matrix $\boldsymbol{\Sigma}_t$.

We construct a portfolio at the beginning of the market and hold it during trading hours. Let $\bar{\mathbf{w}}_d$ be the realized optimal MVP at day d, which satisfies

$$\bar{\mathbf{w}}_d = \arg\min_{\mathbf{h}} \mathbf{h}^\top \boldsymbol{\Gamma}_d \mathbf{h} \quad \text{s.t. } \mathbf{1}^\top \mathbf{h} = 1, \tag{2.2}$$

where $\mathbf{1} = (1, \ldots, 1)^\top \in \mathbb{R}^p$, and $\boldsymbol{\Gamma}_d = \int_{d-1}^d \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^\top dt$ is the daily integrated volatility. Then, we have the following closed form:

$$\bar{\mathbf{w}}_d = \frac{1}{\mathbf{1}^\top \boldsymbol{\Gamma}_d^{-1} \mathbf{1}} \mathbf{w}_d, \quad \text{where} \quad \mathbf{w}_d = \boldsymbol{\Gamma}_d^{-1} \mathbf{1}.$$

The MVP weight vector $\bar{\mathbf{w}}_d$ minimizes the quadratic variation of the portfolio. Specifically, we have the portfolio $\bar{\mathbf{w}}_d^\top \mathbf{X}_t$ during $t \in [d-1, d)$, and its quadratic variation for the dth day is

$$\left[\bar{\mathbf{w}}_d^\top \mathbf{X}_t\right]_d - \left[\bar{\mathbf{w}}_d^\top \mathbf{X}_t\right]_{d-1} = \bar{\mathbf{w}}_d^\top \boldsymbol{\Gamma}_d \bar{\mathbf{w}}_d \quad \text{a.s.}$$

Since $\bar{\mathbf{w}}_d$ satisfies (2.2), the quadratic variation obtains the minimum value among the class

of the weight $\mathbf{h}$ such that $\mathbf{1}^\top \mathbf{h} = 1$. We call $\bar{\mathbf{w}}_d$ the realized MVP.

In this paper, we investigate the realized MVP dynamics and construct dynamic models to accommodate them. We first investigate the non-normalized weight $\mathbf{w}_d = \mathbf{\Gamma}_d^{-1} \mathbf{1}$. We consider the following inverse matrix dynamic model:

$$\mathbf{\Omega}_d = \mathbf{G}_d(\boldsymbol{\theta}) + \mathbf{U}_d, \tag{2.3}$$

where $\mathbf{G}_d(\boldsymbol{\theta})$ is $\mathcal{F}_{d-1}$-adapted, $\boldsymbol{\theta}$ is a model parameter, and $E\left[\mathbf{U}_d | \mathcal{F}_{d-1}\right] = 0$ a.s. Then, we have the following dynamic realized portfolio:

$$\mathbf{w}_d = \mathbf{g}_d(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_d, \tag{2.4}$$

where $\mathbf{g}_d(\boldsymbol{\theta}) = \mathbf{G}_d(\boldsymbol{\theta})\mathbf{1}$, $\boldsymbol{\varepsilon}_d = \mathbf{U}_d\mathbf{1}$, and $\boldsymbol{\varepsilon}_d$'s are martingale differences. Thus, the conditional expected value of $\mathbf{w}_d$ is $E\left[\mathbf{w}_d | \mathcal{F}_{d-1}\right] = \mathbf{g}_d(\boldsymbol{\theta})$ a.s. We call this the dynamic realized minimum variance portfolio (DR-MVP) model. We note that $\mathbf{g}_d(\boldsymbol{\theta})$ can be any function of $\mathcal{F}_{d-1}$-adapted variables, and one of the possible choices is a form of autoregressive (AR), as follows:

$$\mathbf{g}_d(\boldsymbol{\theta}) = \boldsymbol{\beta}_0 + \sum_{j=1}^{q} \boldsymbol{\beta}_j \mathbf{w}_{d-j}, \tag{2.5}$$

where $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ and $\boldsymbol{\beta}_j$'s are p-by-p matrices. There are many structures of $\mathbf{G}_d(\boldsymbol{\theta})$ satisfying (2.4). For example, $\mathbf{G}_d(\boldsymbol{\theta})$ can be the form of BEKK$(0, q)$ (Engle and Kroner, 1995) that can ensure its positive semidefiniteness. Specifically, (2.5) holds if

$$\mathbf{G}_d(\boldsymbol{\theta}) = \mathbf{B}_0 \mathbf{B}_0^\top + \sum_{j=1}^{q} \mathbf{B}_j \mathbf{\Omega}_{d-j} \mathbf{B}_j^\top, \tag{2.6}$$

where $[\mathbf{B}_j]_{kl} = \boldsymbol{\beta}_{j,kl}/\sqrt{\sum_{r=1}^{p} \boldsymbol{\beta}_{j,rl}}$ and $\boldsymbol{\beta}_0 = \mathbf{B}_0 \mathbf{B}_0^\top \mathbf{1}$ under some conditions on $\boldsymbol{\beta}$ or $\mathbf{\Omega}$. The toy examples are as follows. One of the sufficient conditions is that $\mathbf{\Omega}$ is a diagonal

matrix. The diagonal condition allows $\boldsymbol{\beta}$ to be of any form, while the correlation between returns of assets should be zero, which is too restrictive. Another sufficient condition is that all column-sums of $\boldsymbol{\beta}$ are the same. For example, $\boldsymbol{\beta}^\top$ has an eigenvector of $p^{-1/2}\mathbf{1}$. We note that if the corresponding eigenvalue is bounded, then its effect on the inverse matrix is relatively small. Under the $\boldsymbol{\beta}$ restrictive condition, we do not need a condition for the inverse matrix $\boldsymbol{\Omega}$. Alternatively, to relieve the condition, we can consider a compromise condition that balances the diagonal condition and $\boldsymbol{\beta}$ restrictive condition. Specifically, the compromise condition is that if $\sum_{t=1}^{p}\boldsymbol{\beta}_{tk} \neq \sum_{t=1}^{p}\boldsymbol{\beta}_{tl}$, then $\boldsymbol{\Omega}_{kl}$ is zero. Under this condition, (2.5) holds if (2.6) is satisfied. These toy examples illustrate that there exists an inverse matrix dynamic model satisfying the DR-MVP model. In general, we can associate the inverse matrix dynamics model of the BEKK form with the DR-MVP model by dividing the integrated volatility matrix into two parts. One of the matrices follows an inverse matrix dynamic model with $\mathbf{G}_d(\boldsymbol{\theta})$ satisfying (2.6), and the other matrix serves as the supplementary component to make the MVP weights satisfy (2.5). We provide the specific condition in the online Appendix A.1 and utilize it in the simulation study.

We construct a diffusion volatility matrix process that satisfies the inverse matrix dynamic model. In the following theorem, we show the existence of an instantaneous volatility process satisfying (2.3).

Theorem 1. *Suppose that the volatility matrix process satisfies, for any $d \in \mathbb{N}$ and $t \in (d-1, d]$,*

$$\boldsymbol{\Sigma}_t = 2\mathbf{A}_{t-} - \mathbf{A}_{t-}\boldsymbol{\Sigma}_{d-1}^{-1}\mathbf{A}_{t-} + (4\langle t\rangle - 3\langle t\rangle^2)\mathbf{A}_{t-}(\boldsymbol{\Sigma}_{d-1}^{-1} - \mathbf{G}_d(\boldsymbol{\theta}))\mathbf{A}_{t-} - \mathbf{A}_{t-}\mathbf{M}_t\mathbf{A}_{t-}, \quad (2.7)$$

where $\langle t\rangle = t - d + 1$, $\mathbf{A}_t = \langle t\rangle^{-1}\left(\int_{d-1}^{t}\boldsymbol{\Sigma}_s ds\right)$, $\mathbf{M}_t = \int_{d-1}^{t} d\mathbf{Z}_t$, and $\mathbf{Z}_t$ is a symmetric matrix-valued martingale process. Then, its inverse integrated volatility matrices follow the inverse matrix dynamic model (2.3); that is, $\boldsymbol{\Omega}_d = (\int_{d-1}^{d}\boldsymbol{\Sigma}_t dt)^{-1} = \mathbf{G}_d(\boldsymbol{\theta}) + \mathbf{U}_d$, where

7

$\mathbf{U}_d = \int_{d-1}^{d} \mathbf{M}_t dt$ *is a martingale difference.*

Theorem 1 indicates that as long as an instantaneous volatility process satisfies (2.7), its inverse integrated volatility matrix over the dth period can be decomposed into the conditional expectation $\mathbf{G}_d(\boldsymbol{\theta})$ and the martingale difference $\mathbf{U}_d$. Meanwhile, the proposed instantaneous volatility process should satisfy the positive semidefiniteness for any time t. In the online Appendix A.2, we provide sufficient conditions that ensure the positive semidefiniteness of the proposed volatility process. On the other hand, in terms of volatility processes, the proposed instantaneous volatility model accommodates the nonlinear dynamics, which helps capture the MVP dynamics. This is the main difference from the existing parametric high-frequency-based large volatility dynamic models (Kim and Fan, 2019; Shin et al., 2021). The empirical study supports the nonlinear volatility model (see Section 5).

The goal of this paper is to predict the future MVP given the current available information. Thus, the variable of interest is the conditional expected portfolio, namely

$$E\left[\bar{\mathbf{w}}_d|\mathcal{F}_{d-1}\right] = E\left[\left.\frac{\mathbf{w}_d}{\mathbf{1}^\top \mathbf{w}_d}\right|\mathcal{F}_{d-1}\right].$$

Due to the random fluctuation term, this is not exactly the same as the normalized $\mathbf{g}_d(\boldsymbol{\theta})$, $\frac{\mathbf{g}_d(\boldsymbol{\theta})}{\mathbf{1}^\top \mathbf{g}_d(\boldsymbol{\theta})}$. We investigate the error coming from this approximation. To do this, we need the following technical conditions.

Assumption 1.

(a) *For any d, there exist positive constants C_1 and C_2 such that $C_1 \leq p^{-1}\mathbf{1}^\top \mathbf{G}_d(\theta)\mathbf{1} \leq C_2$ and $C_1 \leq p^{-1}\mathbf{1}^\top \boldsymbol{\Omega}_d\mathbf{1} \leq C_2$, where $\mathbf{G}_d(\boldsymbol{\theta}) = E\left[\boldsymbol{\Omega}_d|\mathcal{F}_{d-1}\right].$*

(b) *$\epsilon_{i,d}$'s are sub-Gaussian, and we have $E\left[(\mathbf{1}^\top \boldsymbol{\varepsilon}_d)^2|\mathcal{F}_{d-1}\right] \leq pC$ almost surely, where $\boldsymbol{\varepsilon}_d = (\epsilon_{1,d}, \ldots, \epsilon_{p,d})^\top.$*

Remark 1. Assumption 1(a) is related to the eigenvalue condition for the inverse matrix. That is, if eigenvalues are bounded and strictly bigger than zero, Assumption 1(a) is satisfied. However, when considering factor models, we can have some eigenvalues that converge to zero with p^{-1} order. Furthermore, if $p^{-1/2}\mathbf{1}^\top$ is close to the factor loading matrix space, Assumption 1(a) may not hold. However, in this case, we have that $\mathbf{1}^\top\mathbf{\Omega}_d\mathbf{1}$ has the constant order, and $\mathbf{1}^\top\boldsymbol{\varepsilon}_d$ also has the constant order. Thus, we can obtain the same result in Proposition 1. For simplicity, we impose Assumption 1, and, in this paper, we implicitly assume that the non-normalized weight $\mathbf{w}_d$ is a usual finite random vector that does not depend on p. If it does not hold, by multiplying the corresponding order, we can make the vector have a constant order. Therefore, this condition is not strong. On the other hand, for the inverse matrix, we assume the sparsity (see (3.1)), which implies the sparsity of the random fluctuation part $\mathbf{U}_d$. Thus, it is not restrictive to assume that the row-wise sum of $\mathbf{U}_d$ has a finite random variable (Assumption 1(b)). In fact, if the random fluctuation elements U_{ijd}'s are cross-sectionally independent or weakly dependent, we can obtain Assumption 1(b).

The following proposition shows the error rate of the approximation.

Proposition 1. *Under Assumption 1, we have*

$$\left\| E\left[\bar{\mathbf{w}}_d | \mathcal{F}_{d-1}\right] - \frac{\mathbf{g}_d(\boldsymbol{\theta})}{\mathbf{1}^\top\mathbf{g}_d(\boldsymbol{\theta})} \right\|_{\max} = O_p\left(\left\| \frac{\mathbf{g}_d(\boldsymbol{\theta})}{\mathbf{1}^\top\mathbf{g}_d(\boldsymbol{\theta})} \right\|_{\max} p^{-1/2} + p^{-3/2}\sqrt{\log p} \right), \quad (2.8)$$

$$\left\| E\left[\bar{\mathbf{w}}_d | \mathcal{F}_{d-1}\right] - \frac{\mathbf{g}_d(\boldsymbol{\theta})}{\mathbf{1}^\top\mathbf{g}_d(\boldsymbol{\theta})} \right\|_1 = O_p\left(\left\| \frac{\mathbf{g}_d(\boldsymbol{\theta})}{\mathbf{1}^\top\mathbf{g}_d(\boldsymbol{\theta})} \right\|_1 p^{-1/2} \right). \quad (2.9)$$

Proposition 1 indicates that the conditional expected MVP can be estimated well by the normalized $\mathbf{g}_d(\theta)$ as the number of assets goes to infinity. The error is $p^{-1/2}$ times smaller than the normalized $\mathbf{g}_d(\theta)$, except for the case that the rate of the $\|\mathbf{g}_d(\boldsymbol{\theta})\|_{\max}$ is less than $\sqrt{\log p}$. Thus, we use $\frac{\mathbf{g}_d(\boldsymbol{\theta})}{\mathbf{1}^\top\mathbf{g}_d(\boldsymbol{\theta})}$ as the conditional expected value of the realized MVP. In the following section, we discuss how to estimate $\mathbf{g}_{d-1}(\boldsymbol{\theta})$.

9

3 Estimation Procedure

The intraday log-prices for the dth day are observed at $t_{d,i}, i = 1, \ldots, m_d$, where $d - 1 = t_{d,0} < t_{d,1} < \cdots < t_{d,m_d} = d$. Unfortunately, due to market microstructure noises, true high-frequency observations, $\mathbf{X}_{t_{d,i}}$'s, are not observed. To account for the market microstructure noises, we assume that the observed log-prices $\mathbf{Y}_{t_{d,i}}$ have the following additive noise structure:

$$\mathbf{Y}_{t_{d,i}} = \mathbf{X}_{t_{d,i}} + \mathbf{e}_{t_{d,i}}, \quad \text{for } d = 1, \ldots, N, i = 1, \ldots, m_d,$$

where $\mathbf{X}_t$ is the true log-price and $\mathbf{e}_{t_{d,i}}$'s are microstructure noises with mean zero.

To remove the effect of market microstructure noises, researchers have constructed nonparametric realized volatility estimators that take advantage of subsampling and local averaging techniques so that the integrated volatility can be estimated consistently and efficiently (Aït-Sahalia et al., 2010; Christensen et al., 2010; Fan and Kim, 2018; Jacod et al., 2009; Xiu, 2010; Zhang, 2006, 2011). They demonstrated that the realized volatility estimator has the convergence rate of $m^{-1/4}$, which is known as the optimal convergence rate with the presence of the microstructure noise. In this paper, we employ the pre-averaging realized volatility estimator (Christensen et al., 2010; Jacod et al., 2009) in the numerical study. We denote the realized volatility matrix estimator by $\widehat{\boldsymbol{\Gamma}}$.

3.1 Non-parametric Realized Minimum Variance Portfolio Estimator

We first introduce some notations. For any given p_1 by p_2 matrix $\mathbf{M} = (M_{ij})$, the Frobenius norm of $\mathbf{M}$ is denoted by $\|\mathbf{M}\|_F = \sqrt{\text{tr}(\mathbf{M}^\top \mathbf{M})}$, the matrix spectral norm $\|\mathbf{M}\|_2$ is the

square root of the largest eigenvalue of $\mathbf{M}\mathbf{M}^\top$, and let

$$\|\mathbf{M}\|_1 = \max_{1 \leq j \leq p_2} \sum_{i=1}^{p_1} |M_{ij}|, \quad \|\mathbf{M}\|_\infty = \max_{1 \leq i \leq p_1} \sum_{j=1}^{p_2} |M_{ij}|, \quad \|\mathbf{M}\|_{\max} = \max_{i,j} |M_{ij}|.$$

C's denote generic positive constants whose values are free of other parameters and may change from appearance to appearance.

In this section, we introduce a realized MVP estimation procedure. To estimate the realized MVP, we first estimate the inverse matrix of the integrated volatility matrix. We assume that the inverse matrix of the integrated volatility matrix, $\mathbf{\Gamma}_d^{-1} = \mathbf{\Omega}_d = (\Omega_{ij,d})_{i,j=1,\dots,p}$, satisfies the following sparsity condition:

$$\sup_d \max_{1 \leq i \leq p} \sum_{j=1}^{p} |\Omega_{ij,d}|^\delta \leq s_p \text{ a.s.}, \tag{3.1}$$

where $\delta \in [0, 1)$ and s_p is diverging slowly with respect to p, such as $\log p$. To accommodate this sparsity, we use the constrained ℓ_1-minimization for inverse matrix estimation (CLIME) (Cai et al., 2011) with the realized volatility matrix estimator. Specifically, let $\widehat{\mathbf{\Omega}}_d$ be the solution for the following optimization problem:

$$\min \|\mathbf{A}\|_1 \quad \text{s.t.} \quad \|\widehat{\mathbf{\Gamma}}_d \mathbf{A} - \mathbf{I}\|_{\max} \leq \tau_m, \tag{3.2}$$

where τ_m is the tuning parameter specified in Proposition 2. Then, we estimate the realized non-normalized MVP estimator as $\widehat{\mathbf{w}}_d = \widehat{\mathbf{\Omega}}_d \mathbf{1}$. To investigate its asymptotic behavior, we need the following technical conditions.

Assumption 2.

(a) There exists a large C_a depending on given constant a such that

$$P\left(|\widehat{\Gamma}_{ij,d} - \Gamma_{ij,d}| \geq C_a \sqrt{\log(p \vee N)} m^{-1/4}\right) \leq C(p \vee N)^a.$$

(b) $\max_{d \leq N} \|\mathbf{\Omega}_d\|_1 \leq C$ *a.s.*

Remark 2. The sub-exponential tail condition in Assumption 2(a) is often imposed to handle high-dimensional statistics. Under the bounded instantaneous and drift condition, we can obtain the sub-exponential condition (Kim and Wang, 2016; Tao et al., 2013). In contrast, for heavy-tailed observations, by employing the truncation method, we can obtain Assumption 2(a) (Fan and Kim, 2018; Shin et al., 2023). Thus, Assumption 2(a) is not restrictive.

The following proposition establishes the convergence rates for the inverse matrix estimator and realized non-normalized MVP estimator.

Proposition 2. *Under Assumption 2 and the sparsity condition (3.1), we choose $\tau_m = C_\tau m^{-1/4} \sqrt{\log(p \vee N)}$ for some large constant C_τ. Then, we have*

$$\max_{d \leq N} \|\widehat{\mathbf{\Omega}}_d - \mathbf{\Omega}_d\|_{\max} \leq C\tau_m, \qquad \max_{d \leq N} \|\widehat{\mathbf{\Omega}}_d - \mathbf{\Omega}_d\|_1 \leq C s_p \tau_m^{1-\delta}, \quad and \qquad (3.3)$$

$$\max_{d \leq N} \|\widehat{\mathbf{w}}_d - \mathbf{w}_d\|_{\max} \leq C s_p \tau_m^{1-\delta} \qquad (3.4)$$

with probability greater than $1 - (p \vee N)^{-c}$ for any given positive constant c.

Remark 3. Proposition 2 indicates that the CLIME estimator has the convergence rate $m^{-1/4} \sqrt{\log(p \vee N)}$. When microstructure noise is present, it is known that $m^{-1/4}$ is the optimal rate (Tao et al., 2013). Thus, the CLIME estimator can obtain the optimal rate up to log order. On the other hand, since the realized non-normalized MVP estimator $\widehat{\mathbf{w}}_d$ is the form of the row-wise sum of the CLIME estimator, the convergence rate of $\widehat{\mathbf{w}}_d$ is bounded by the ℓ_1 norm bound of the CLIME estimator. When we consider the exact sparsity ($\delta = 0$), $\widehat{\mathbf{w}}_d$ has the convergence rate $s_p m^{-1/4} \sqrt{\log(p \vee N)}$. The term s_p is the number of non-zero elements, which is usually assumed to be negligible in high-dimensional statistics, such as $\log p$. Thus, $\widehat{\mathbf{w}}_d$ has the optimal rate $m^{-1/4}$ up to log order.

3.2 Model Parameter Estimator

In this paper, we assume that $\mathbf{g}_d(\boldsymbol{\theta})$ is a function of the past realized non-normalized MVP weights $\mathbf{w}_{d-1}, \ldots, \mathbf{w}_1$. That is, we consider a time series structure. We note that it would be straightforward to extend this to a more general model by including exogenous variables. We estimate $\mathbf{g}_d(\boldsymbol{\theta})$ by using the plug-in method. Specifically, we use the realized non-normalized MVP estimator $\widehat{\mathbf{w}}_d$'s instead of $\mathbf{w}_d$'s. We denote this estimator by $\widehat{\mathbf{g}}_d(\boldsymbol{\theta})$. For example, when we consider the $\mathrm{AR}(q)$ structure, we have

$$\widehat{\mathbf{g}}_d(\boldsymbol{\theta}) = (\widehat{g}_{i,d}(\boldsymbol{\theta}_1), \ldots, \widehat{g}_{p,d}(\boldsymbol{\theta}_p))^\top = \boldsymbol{\beta}_0 + \sum_{j=1}^{q} \boldsymbol{\beta}_j \widehat{\mathbf{w}}_{d-j}.$$

In this section, since the empirical study supports the AR structure (see Figure 1), we derive asymptotic theorems based on the $\mathrm{AR}(q)$ model for simplicity. However, for any well-defined $\mathbf{g}_d(\theta)$, we can derive similar results under some regularity conditions.

The model is based on the high-dimensional vector autoregressive (VAR) model. Thus, we suffer from the curse of dimensionality. To tackle this obstacle, we assume that the coefficient $\boldsymbol{\beta}_j$ is sparse (see Assumption 3(b)). To accommodate the sparsity, we employ the LASSO for each weight as follows:

$$\widehat{\boldsymbol{\theta}}_i = \arg\min_{\boldsymbol{\theta}_i} \mathcal{L}_{n,i}(\boldsymbol{\theta}_i) + \lambda_n \|\boldsymbol{\theta}_i\|_1,$$

where $\mathcal{L}_{n,i}(\boldsymbol{\theta}_i) = \frac{1}{n}\sum_{d=1}^{n}\left(\widehat{w}_{i,q+d} - \widehat{g}_{i,q+d}(\boldsymbol{\theta}_i)\right)^2$ with $n = N - q$ samples and λ_n is a tuning parameter defined in Theorem 2. We note that the lag order q is allowed to increase with N. To analyze its asymptotic behaviors, we need the following technical conditions.

Assumption 3.

(a) $\mathbf{W}_d = (\mathbf{w}_d^\top, \ldots, \mathbf{w}_{d-q+1}^\top)^\top$ *is strictly stationary, and the spectral radius of* $\mathbf{A}$ *is less*

than one, where

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_{q-1} & \boldsymbol{\beta}_q \\ \mathbf{I}_p & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{I}_p & 0 & \cdots & 0 \\ \vdots & & \ddots & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}_p & 0 \end{pmatrix}.$$

(b) *The number of nonzero elements of $\boldsymbol{\theta}_{0,i}$ is bounded by $s_\beta \geq 1$, where $\boldsymbol{\theta}_{0,i}$ is the true parameter for the ith weight.*

(c) *The process $\mathbf{W}_d$ is α-mixing, and the α mixing coefficients satisfy $\alpha(k) = O(c^k)$ for some $c \in (0, 1)$.*

(d) $\max_d \|\widehat{\mathbf{w}}_d - \mathbf{w}_d\|_{\max} \leq C s_p \tau_m^{1-\delta}$.

(e) *Define $\mathcal{U}_i = \left\{ \mathbf{u} \in \mathbb{R}^{qp+1} : \|\mathbf{u}_{S_i^c}\|_1 \leq 3 \|\mathbf{u}_{S_i}\|_1 \right\}$, where $S_i = \{j : jth$ element of $\boldsymbol{\theta}_{0,i} \neq 0\}$, $\mathbf{u}_{S_i}$ is the subvector obtained by stacking $\{\mathbf{u}_j : j \in S_i\}$, and $\mathbf{u}_{S_i^c}$ is the subvector obtained by stacking $\{\mathbf{u}_j : j \in S_i^c\}$. Then, there exists a constant $\kappa > 0$ such that the following inequality holds for some $D \geq 48 s_\beta \lambda_n / \kappa$ and $1 \leq i \leq p$, where the specific value of λ_n is given in Theorem 2:*

$$\inf\{\mathbf{u}^\top \nabla^2 \mathcal{L}_{n,i}(\boldsymbol{\theta}_i)\mathbf{u} : \ \mathbf{u} \in \mathcal{U}_i, \ \|\mathbf{u}\|_2 = 1, \ \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0,i}\|_1 \leq D\} \geq \kappa.$$

Remark 4. Assumption 3(a) is the strictly stationary and stable condition for the VAR(1) model for $\mathbf{W}_d$, which is a first-order representation of (2.5). For the model (2.7), the strictly stationarity holds if $\mathbf{Z}_t$ is a strictly stationary process and the spectral radius of $\mathbf{A}$ is less than one. Assumption 3(c) is satisfied with an exponentially decaying absolute summation condition on coefficients of MA(∞) representations of $\mathbf{W}_d$ (see Assumption (A2) in Masini et al. (2022) and Theorem 2.1 in Chanda (1974)), which restricts the coefficients $\boldsymbol{\beta}_j$'s.

Assumption 3(d) is satisfied with high probability by Proposition 2. In contrast, when we employ other estimation procedures for the MVP weights, as long as Assumption 3(d)-type condition is satisfied, we can obtain a similar result. That is, as long as the realized non-normalized MVP estimator performs well, the proposed estimation method works. Finally, Assumption 3(e) is the eigenvalue conditions for the Hessian matrix $\nabla^2 \mathcal{L}_{n,i}(\boldsymbol{\theta}_i)$. This is called the localized restricted eigenvalue (LRE) condition (Fan et al., 2018; Shin et al., 2021; Sun et al., 2020), which implies strictly positive restricted eigenvalues over a local neighborhood.

The following theorem establishes the asymptotic convergence rate for the LASSO estimation procedure.

Theorem 2. *Suppose that N is greater than the lag order q such that the number of usable observations $n = N - q$ diverges as N increases. Under Assumptions 1–3, take $\lambda_n \geq C(s_\beta s_p \tau_m^{1-\delta} + n^{-1/2}\sqrt{\log(p \vee q)})$ for some large constant C, which goes to zero as p and n go to infinity. Then, we have*

$$\max_i \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{0,i}\|_1 \leq C\frac{s_\beta \lambda_n}{\kappa} \quad and \quad \max_i \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{0,i}\|_2 \leq C\frac{s_\beta^{1/2} \lambda_n}{\kappa} \tag{3.5}$$

with probability at least $1 - (p \vee q)^{-c}$ for any given positive constant c.

Remark 5. Theorem 2 shows that the LASSO estimator has the convergence rate λ_n with s_β term. When we consider exact sparsity for the inverse matrix ($\delta = 0$), we have the convergence rate $m^{-1/4} + n^{-1/2}$ with additional s_p and s_β up to log order. The term $m^{-1/4}$ is the cost to estimate the unobserved non-normalized MVP using the high-frequency observations. The term $n^{-1/2}$ is the usual convergence rate for estimating low-frequency models with the n period. Given the lag order q, the total number of variables for the LASSO estimator is pq. Therefore, the lag order q and the number of assets p affect the convergence rate λ_n of the LASSO estimator.

The main purpose of this paper is to predict the future MVP. To do this, we estimate the conditional MVP weight as $\widehat{\mathbf{w}}_{n+1} = \frac{\widehat{\mathbf{g}}_{n+1}(\widehat{\boldsymbol{\theta}}_i)}{\mathbf{1}^\top \widehat{\mathbf{g}}_{n+1}(\widehat{\boldsymbol{\theta}}_i)}$. In the following theorem, we investigate its asymptotic properties.

Theorem 3. *Under the assumptions in Theorem 2, we have*

$$\|\widehat{\mathbf{g}}_{n+1}(\widehat{\boldsymbol{\theta}}) - \mathbf{g}_{n+1}(\boldsymbol{\theta}_0)\|_{\max} \leq C \frac{s_\beta \lambda_n}{\kappa} \tag{3.6}$$

and

$$\|\widehat{\mathbf{w}}_{n+1} - E\left(\bar{\mathbf{w}}_{n+1} | \mathcal{F}_n\right)\|_1 \leq C \left(\frac{s_\beta \lambda_n}{\kappa} + \left\| \frac{\mathbf{g}_{n+1}(\boldsymbol{\theta})}{\mathbf{1}^\top \mathbf{g}_{n+1}(\boldsymbol{\theta})} \right\|_1 p^{-1/2} \right) \tag{3.7}$$

with probability at least $1 - (p \vee q)^{-c}$ for any given positive constant c.

Theorem 3 indicates that the proposed conditional expected non-normalized MVP estimator $\widehat{\mathbf{g}}_{n+1}(\widehat{\boldsymbol{\theta}})$ can consistently estimate the future non-normalized MVP with the $s_\beta \lambda_n$ order. Furthermore, the conditional MVP weight estimator has the convergence rate $s_\beta \lambda_n + \left\| \frac{\mathbf{g}_{n+1}(\boldsymbol{\theta})}{\mathbf{1}^\top \mathbf{g}_{n+1}(\boldsymbol{\theta})} \right\|_1 p^{-1/2}$. The first term $s_\beta \lambda_n$ is the cost to estimate unobserved future non-normalized MVP using the high-frequency and low-frequency observations. The second term $\left\| \frac{\mathbf{g}_{n+1}(\boldsymbol{\theta})}{\mathbf{1}^\top \mathbf{g}_{n+1}(\boldsymbol{\theta})} \right\|_1 p^{-1/2}$ is the approximation error. From Theorem 3, we can conclude that the proposed estimation method can consistently estimate the future MVP.

3.3   Discussion on Price Jumps

In financial practice, we often observe jumps. In the presence of price jumps, quadratic covariation can be decomposed into continuous and jump components, and the product of its inverse and one vector is the non-normalized MVP in the ex-post sense. On the other hand, several empirical studies reported that the volatility dynamics can be better explained by decomposing quadratic covariation into its continuous and jump components, since jumps are associated with news announcements and have less persistent features (Andersen

et al., 2007; Barndorff-Nielsen and Shephard, 2006; Corsi et al., 2010; Lee and Mykland, 2008). In a similar spirit, employing MVP using integrated volatility rather than quadratic covariation may help capture the conditional MVP dynamics well. To check this, we conduct an empirical study to compare integrated volatility and quadratic covariation and report the results in the online Appendix D. We find that employing integrated volatility shows better performance. Details can be found in the online Appendix D.

To estimate integrated volatility in the presence of price jumps, we first detect jumps from observed stock log-return data and estimate the integrated volatility. For example, Fan and Wang (2007) and Zhang et al. (2016) employed the wavelet method to identify the jumps based on noisy high-frequency data. Mancini (2004) studied a threshold method for jump detection and presented the order of an optimal threshold, and Davies and Tauchen (2018) further examined a data-driven threshold method. They demonstrated that the estimator of jump variation has the convergence rate of $m^{-1/4}$, which further helps the estimator of integrated volatility to achieve the optimal convergence rate of $m^{-1/4}$. In this paper, we employ the jump-robust pre-averaging realized volatility matrix estimator (JPRVM) (Aït-Sahalia and Xiu, 2016) and describe the procedure in the online Appendix B. Then, we can show that this estimator satisfies Assumption 2(a) under the sub-Gaussian condition, and with it, we can apply the proposed estimation procedure and obtain the same result.

3.4 Choice of the Tuning Parameters

To apply the proposed estimators, we need to choose the tuning parameters τ_m and λ_n for the CLIME estimator and the LASSO estimator, respectively. For the CLIME estimator, we estimate it by varying C_τ of $\tau_m = C_\tau m^{-1/4}\sqrt{\log(p \vee N)}$, and we select C_τ among 100 logarithmically spaced points ranging from 10^{-6} to 10, which minimizes the likelihood loss,

as follows:

$$\widehat{\Omega}_{d,\tau} = \mathrm{argmin}_{\Omega} \, \|\Omega\|_1 \ \text{ s.t. } \ \left\|\widehat{\Gamma}_d \Omega - \mathbf{I}\right\|_{\max} \leq \tau,$$

$$\widehat{\Omega}_d = \widehat{\Omega}_{d,\tau^*}, \text{ where } \tau^* = \underset{\tau \in \{\tau_m | C_\tau \in [10^{-6}, 10]\}}{\mathrm{argmin}} \langle \widehat{\Omega}_{d,\tau}, \widehat{\Gamma}_d \rangle - \log \det(\widehat{\Omega}_{d,\tau}).$$

For the LASSO estimator, we need to choose the tuning parameters λ_n. The estimation of the DR-MVP model typically faces the small-n-large-P situation (Chen and Chen, 2008). For example, in our empirical study (Section 5), we have $n = 252 - 50 = 202$ and $P = 200 \times 50 = 10000$ for the estimation of the DR-MVP model with an AR(50) specification. Therefore, we select the tuning parameter λ_n from 100 logarithmically spaced points ranging from 10^{-6} to 10, which minimizes the extended Bayesian information criterion (EBIC) (Chen and Chen, 2008, 2012) of the LASSO estimation, where

$$\mathrm{EBIC} = n \log(\mathcal{L}_{n,i}(\boldsymbol{\theta}_i)) + \|\boldsymbol{\theta}_i\|_0 \log(n) + 2 \|\boldsymbol{\theta}_i\|_0 \gamma \log P,$$

$\mathcal{L}_{n,i} = \frac{1}{n} \sum_{d=1}^n (\widehat{w}_{i,d} - \widehat{g}_{i,d}(\boldsymbol{\theta}_i))^2$, $\gamma = 0.5$, as Chen and Chen (2012) suggested, and P represents the number of covariates under consideration.

4 Simulation Study

We conducted simulations to show that the proposed methodology has good finite sample performance and compared the proposed method with other existing methods. In Section 5, we find that the heterogeneous autoregressive (HAR) specification for the DR-MVP model can help explain the MVP dynamics. To reflect this feature, we considered the HAR-type VAR(22) structure for the DR-MVP model, which is

$$\mathbf{w}_d = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \bar{\mathbf{w}}_d^{(1)} + \boldsymbol{\beta}_5 \bar{\mathbf{w}}_d^{(5)} + \boldsymbol{\beta}_{22} \bar{\mathbf{w}}_d^{(22)} + \boldsymbol{\varepsilon}_d, \tag{4.1}$$

where $\bar{\mathbf{w}}_d^{(k)} = k^{-1} \sum_{j=1}^k \mathbf{w}_{d-j}$ is the average level of MVP weights over the previous k days. The true log-prices for the p assets follow the jump-diffusion process

$$dX_t = \boldsymbol{\sigma}_t d\mathbf{B}_t + \mathbf{J}_t d\boldsymbol{\Lambda}_t,$$

where $\mathbf{B}_t$ is a p-dimensional standard Brownian motion, $\boldsymbol{\Sigma}_t = \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^\top$ is the instantaneous volatility process, the jump sizes $\mathbf{J}_t$ obey the independent normal distribution $N(0.05, 0.005^2)$ with randomly determined signs, and $\boldsymbol{\Lambda}_t$ is a p-dimensional Poisson process with intensity 5. The instantaneous volatility matrix $\boldsymbol{\Sigma}_t$ follows the data-generating process introduced in the online Appendix A, so the DR-MVP model in (4.1) is satisfied. For each simulation process, we generated high-frequency data with $m = 23400$ for 500 consecutive days. We used the subsampled log-prices of the last $N = 125, 250, 500$ days with high-frequency observations $m = 2340, 7800, 23400$ per day and repeated the simulation procedure 100 times.

To estimate the integrated volatility matrix, we utilized the POET procedure (Fan et al., 2013). Specifically, the input integrated volatility matrix for the POET procedure is estimated by the JPRVM estimator in the online Appendix B. We then took a hard thresholding function with threshold level $\sqrt{\log p / m^{1/2}} + 1/\sqrt{p}$ for the idiosyncratic volatility matrix.

Figure 2 draws the mean of matrix ℓ_1-norms and max norms of $\widehat{\boldsymbol{\Omega}}_d - \boldsymbol{\Omega}_d$ and mean max and ℓ_1-norms of $\widehat{\mathbf{w}}_d - \mathbf{w}_d$ for the CLIME and realized non-normalized MVP estimators with $m = 2340, 7800, 23400$. From Figure 2, we find that the mean estimation errors of non-parametric estimators decrease as the number of high-frequency observations increases. These results support the theoretical results derived in Proposition 2.

We then checked the finite sample performances of the DR-MVP model. We first estimated the model parameters by the LASSO for $N = 125, 250, 500$ and $m = 2340, 7800, 23400$,
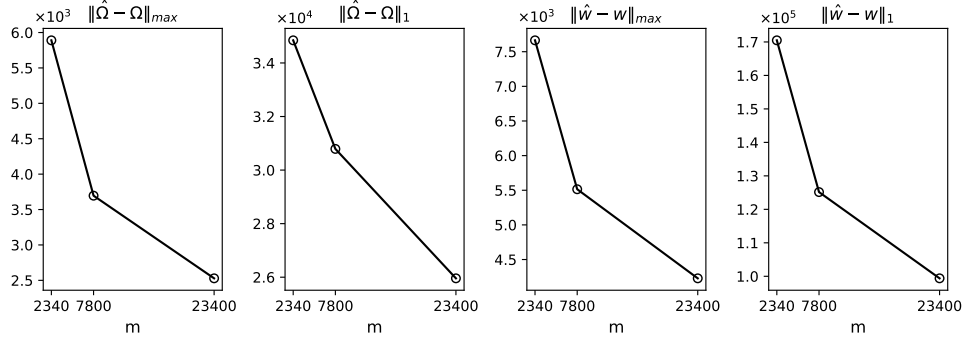
Figure 2: Mean of matrix max and ℓ_1-norms of $\widehat{\boldsymbol{\Omega}}_d - \boldsymbol{\Omega}_d$ and mean max and ℓ_1-norms of $\widehat{\mathbf{w}}_d - \mathbf{w}_d$ for the CLIME and realized non-normalized MVP estimators with $m = 2340, 7800, 23400$.

and the cases employing the true MVPs. Figure 3 depicts the mean errors of $\max_i \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{0,i}\|_1$ and $\max_i \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{0,i}\|_2$ for the DR-MVP estimator. From Figure 3, we find that the mean errors decrease as the number of high- or low-frequency observations increases. These results support the theoretical findings in Theorem 2.
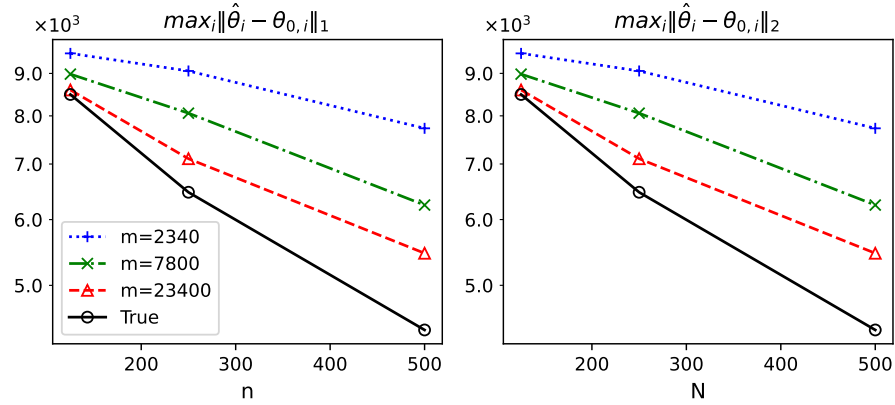

Figure 3: Mean errors of $\max_i \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{0,i}\|_1$ and $\max_i \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{0,i}\|_2$ for the proposed DR-MVP estimator with $N = 125, 250, 500$ and $m = 2340, 7800, 23400$, and the true MVP.

The main purpose of this paper is to predict the future MVP. Thus, we investigated the out-of-sample performance of estimating the one-day-ahead conditional non-normalized and normalized MVP $\mathbf{g}_{d+1}(\boldsymbol{\theta}_0)$ and $\mathbb{E}\left[\bar{\mathbf{w}}_{d+1}|\mathcal{F}_d\right]$, where we calculated $\mathbb{E}\left[\bar{\mathbf{w}}_{d+1}|\mathcal{F}_d\right]$ by the Monte Carlo method. For comparisons, we considered one additional DR-MVP model that utilizes the simple HAR model (HAR) (Corsi, 2009) for estimating the future non-

normalized MVP weights, and HAR employs the OLS procedure to estimate parameters. We also consider a non-parametric estimator that only utilizes the latest CLIME estimator, such as $\widehat{\boldsymbol{\Omega}}_d \mathbf{1}$ and $\frac{\widehat{\boldsymbol{\Omega}}_d \mathbf{1}}{\mathbf{1}^\top \widehat{\boldsymbol{\Omega}}_d \mathbf{1}}$. The CLIME estimator is based on the martingale assumption for the non-normalized MVP weights. Finally, we employed the parametric estimator FIVAR (Shin et al., 2021), which can account for the factor and idiosyncratic volatility dynamics. With regard to the FIVAR estimator, we first obtained the estimated conditional volatility matrix $\widehat{\boldsymbol{\Gamma}}_{d+1}^F$ by the FIVAR estimator, and we calculated the non-normalized and normalized MVPs as $(\widehat{\boldsymbol{\Gamma}}_{d+1}^F)^{-1}\mathbf{1}$ and $\frac{(\widehat{\boldsymbol{\Gamma}}_{d+1}^F)^{-1}\mathbf{1}}{\mathbf{1}^\top (\widehat{\boldsymbol{\Gamma}}_{d+1}^F)^{-1}\mathbf{1}}$, respectively. To estimate the one-day-ahead FIVAR volatility matrix, we followed the same procedure suggested in Shin et al. (2021).
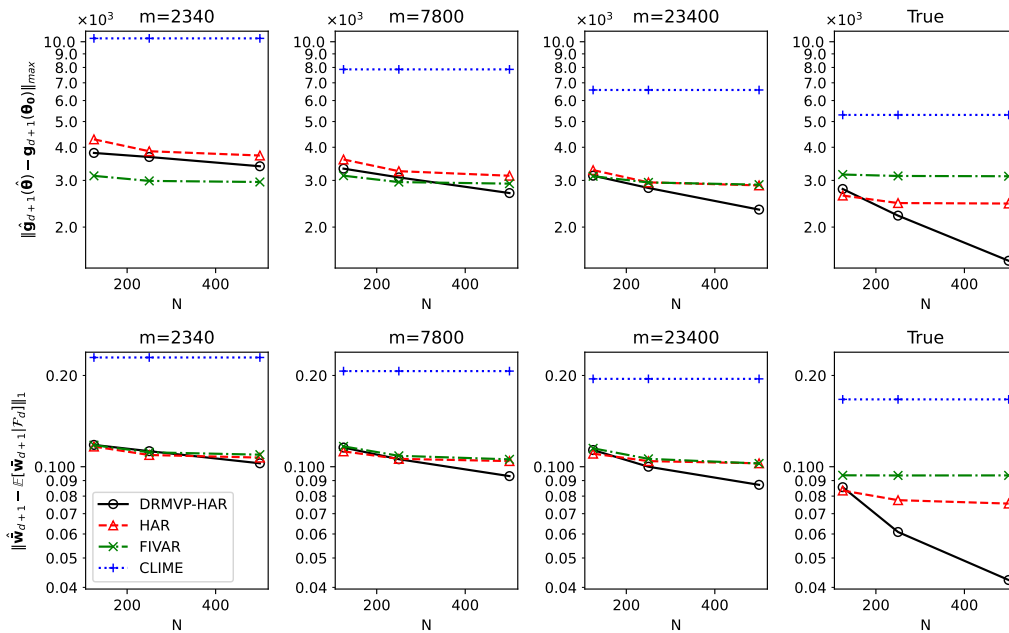

Figure 4: Mean of $\|\widehat{\mathbf{g}}_{d+1}(\widehat{\boldsymbol{\theta}}) - \mathbf{g}_{d+1}(\boldsymbol{\theta}_0)\|_{\max}$ and $\|\widehat{\bar{\mathbf{w}}}_{d+1} - \mathbb{E}[\bar{\mathbf{w}}_{d+1}|\mathcal{F}_d]\|_1$ for $N = 125, 250, 500$ and $m = 2340, 7800, 23400$, along with cases employing the true values in place of nonparametric estimators.

To provide numerical support for Theorem 3, Figure 4 draws the mean of $\|\widehat{\mathbf{g}}_{d+1}(\widehat{\boldsymbol{\theta}}) - \mathbf{g}_{d+1}(\boldsymbol{\theta}_0)\|_{\max}$ and $\|\widehat{\bar{\mathbf{w}}}_{d+1} - \mathbb{E}[\bar{\mathbf{w}}_{d+1}|\mathcal{F}_d]\|_1$ for $N = 125, 250, 500$ and $m = 2340, 7800, 23400$, along with cases employing the true values in place of nonparametric estimators. From Figure 4, we find that the parametric models perform better than the nonparametric model

CLIME. When comparing the parametric models, the DRMVP-HAR model performs better than others for higher m and n, while the FIVAR model performs better in terms of $\|\widehat{\mathbf{g}}_{d+1}(\widehat{\boldsymbol{\theta}}) - \mathbf{g}_{d+1}(\boldsymbol{\theta}_0)\|_{\max}$ for $m = 2340$. This implies that we need sufficient high-frequency observations to utilize the estimated MVP as inputs of the DR-MVP models. Moreover, we find that the mean prediction errors of the conditional non-normalized and normalized MVPs decrease as the number of high- or low-frequency observations increases. These results support the theoretical results in Theorem 3. We note that additional prediction error analyses for different norms can be found in the online Appendix C.

We further investigated the out-of-sample risk of the minimum variance portfolio allocation problem. The out-of-sample risk of a normalized portfolio $\widehat{\overline{\mathbf{w}}}_{d+1}$ is calculated as $R = \widehat{\overline{\mathbf{w}}}_{d+1}^{\top} \boldsymbol{\Gamma}_{d+1} \widehat{\overline{\mathbf{w}}}_{d+1}$, where $\boldsymbol{\Gamma}_{d+1}$ is the one-day-ahead integrated volatility matrix. Figure 5 depicts the mean out-of-sample normalized portfolio risk for the DR-MVP, FIVAR, and CLIME models with $N = 125, 250, 500$ and $m = 2340, 7800, 23400$, and the cases employing the true values in place of nonparametric estimators. The horizontal black solid line in Figure 5 indicates the mean out-of-sample normalized portfolio risk that can be achieved with the true $\mathbf{g}_{d+1}(\boldsymbol{\theta})$. Figure 5 shows that the DR-MVP model performs best among the benchmarks, except for the small m case ($m = 2340$). This is because the DR-MVP model can directly capture the MVP dynamics. On the other hand, the mean out-of-sample risk for the DR-MVP, FIVAR, and CLIME models decreases as the number of high- or low-frequency observations increases. This may be because, as the number of high- or low-frequency observations increases, we can obtain more accurate input volatility matrix estimator and parametric estimations, respectively.

We end this section by remarking that the proposed DR-MVP model shows comparable performance when the true model is misspecified. To assess the robustness of the proposed methodology when the true model is misspecified, we conducted an additional analysis assuming that the data-generating process follows a different specification, namely the
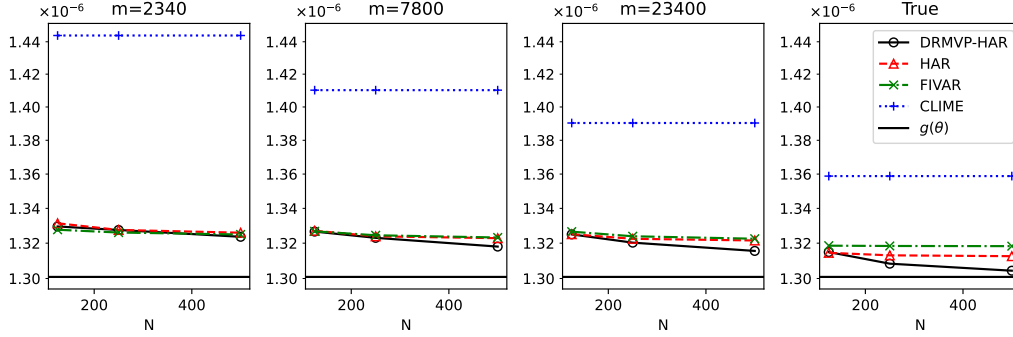
Figure 5: Mean out-of-sample normalized portfolio risk for the DR-MVP, FIVAR, and CLIME models with $N = 125, 250, 500$ and $m = 2340, 7800, 23400$, along with cases employing the true values in place of nonparametric estimators.

FIVAR data generating process (Shin et al., 2021). The full methodology and results of this robustness analysis are presented in the online Appendix B.

5 Empirical study

In this section, we apply the proposed DR-MVP model to real high-frequency stock trading data. We obtained high-frequency data from January 2016 to December 2017, 503 trading days in total, for the top 200 large trading volume stocks among the components of the S&P 500 from the TAQ database in the Wharton Research Data Service (WRDS) system. In order to calculate the covariances between different assets, we employed the pairwise-refresh time scheme proposed by Fan et al. (2012) to obtain synchronized samples. The average values of daily means, medians, standard deviations, minimums, and maximums of the number of daily synchronized samples are 6283, 5474, 4030, 1968, and 225180, respectively.

To employ the proposed DR-MVP model, we first needed to obtain the inverse matrix estimators of the realized volatility matrix estimators by the CLIME estimation procedure. For the input volatility matrix, we used the JPRVM estimator defined in the online Appendix B to the synchronized intraday log-prices for each pair of stocks. We then employed the POET procedure introduced by Fan et al. (2013, 2016) and used the global industry

classification standard (GICS) for the threshold level for the idiosyncratic part (Fan et al., 2016). Specifically, we kept the idiosyncratic volatilities within the same sector, but set zero for others. Then, with the POET estimator, we applied the CLIME estimator to obtain the inverse matrix estimations $\widehat{\Omega}_d$ and the MVPs $\widehat{\mathbf{w}}_d = \widehat{\Omega}_d \mathbf{1}$ with the tuning parameters τ_n, chosen by the procedure in Section 3.4.

We then estimated $\widehat{\mathbf{g}}_{d+1}(\widehat{\boldsymbol{\theta}})$ by LASSO with the tuning parameters λ_n, chosen by the procedure in Section 3.4. For the explanatory variables of $\widehat{\mathbf{g}}_{d+1}(\boldsymbol{\theta})$, we considered two cases, such as 50 lagged realized MVPs (DRMVP-50) and HAR-type realized MVPs (DRMVP-HAR). The latter can reproduce relatively long memory persistence, such as AR(22), while remaining parsimonious. Specifically, we assumed the VAR structure for $\widehat{\mathbf{g}}_{d+1}(\boldsymbol{\theta})$ with HAR-type realized MVPs as follows:

$$\widehat{\mathbf{g}}_{d+1}(\boldsymbol{\theta}) = \boldsymbol{\beta}_0 + \sum_{h \in \{1,5,22\}} \boldsymbol{\beta}_h \widehat{\mathbf{w}}_d^{(h)}, \tag{5.1}$$

where $\widehat{\mathbf{w}}_d^{(h)} = \frac{1}{h} \sum_{j=d-h+1}^{d} \widehat{\mathbf{w}}_j$. We used the rolling window scheme to predict the one-day-ahead non-normalized MVPs, where the in-sample period was 252 days.

To check if the proposed model had captured the MVP dynamics well, we investigated the out-of-sample performance of the minimum variance portfolio allocation problem. We computed the out-of-sample annualized portfolio risk of the normalized portfolios for a given period with d days. Furthermore, to assess how close the out-of-sample portfolio risk is to the ex-post minimum variance portfolio risk, we calculated the mean relative portfolio risks. Specifically, the annualized risk and the mean relative risk are calculated as follows:

$$\text{Annualized risk} = \sqrt{\frac{252}{d} \sum_{k=1}^{d} \sum_{i=1}^{39} r_{k,i}(\widehat{\mathbf{w}}_k)^2} \quad \text{and Mean relative risk} = \frac{1}{d} \sum_{k=1}^{d} \frac{\sum_{i=1}^{39} r_{k,i}(\widehat{\mathbf{w}}_k)^2}{\sum_{i=1}^{39} r_{k,i}(\bar{\mathbf{w}}_k)^2},$$

where $\widehat{\mathbf{w}}_k$ is a normalized portfolio and $r_{k,i}(\widehat{\mathbf{w}}_k) = \widehat{\mathbf{w}}_k^\top (\mathbf{Y}(k-1+\frac{i}{39}) - \mathbf{Y}(k-1+\frac{i-1}{39}))$

24

is the 10-minute portfolio log-return and $\bar{\mathbf{w}}_k$ is the ex-post normalized minimum variance portfolio obtained from the CLIME estimator. Finally, we computed the mean ℓ_2-norm of the difference between the estimated out-of-sample normalized MVP and the ex-post normalized MVP, $\sum_{k=1}^{d} \|\widehat{\mathbf{w}}_k - \bar{\mathbf{w}}_k\|_2/d$. We used two different out-of-sample periods, namely from day 253 to day 378 and from day 379 to day 503, denoted by Period 1 and Period 2, respectively. We considered two benchmarks based on the CLIME and FIVAR estimators and one additional DR-MVP model that utilizes the simple HAR model (HAR), defined in Section 4. For example, HAR follows the proposed DR-MVP model, but the off-diagonal elements of the $(\boldsymbol{\beta}_{i,h})_{h=1,5,22}$ are restricted to zero and $\widehat{\boldsymbol{\theta}}_i = \arg\min_{\boldsymbol{\theta}_i} \mathcal{L}_{n,i}(\boldsymbol{\theta}_i)$. All the benchmark portfolios were normalized to measure the out-of-sample portfolio risks.

Table 1: Out-of-sample annualized and relative risks of the normalized MVP with the DR-MVP, HAR, FIVAR, and CLIME models for Period 1, Period 2, and the whole period.

	Annualized risks				
Period	DRMVP-50	DRMVP-HAR	HAR	FIVAR	CLIME
1	4.667%	4.660%	4.512%	4.923%	4.801%
2	4.199%	4.182%	4.124%	4.447%	4.333%
Whole	4.438%	4.427%	4.322%	4.690%	4.572%
	Mean relative risks				
Period	DRMVP-50	DRMVP-HAR	HAR	FIVAR	CLIME
1	1.339	1.337	1.301	1.434	1.379
2	1.358	1.351	1.323	1.479	1.412
Whole	1.348	1.344	1.312	1.457	1.396

Table 1 reports the out-of-sample annualized risks and relative risks of the normalized MVP with the DRMVP-50, DRMVP-HAR, HAR, FIVAR, and CLIME models. From Table 1, we find that the DR-MVP-based models (DRMVP-50, DRMVP-HAR, and HAR) show the best performance. The HAR model is slightly better than the DRMVP-50 and DRMVP-HAR models. This may be because the DR-MVP models help capture the MVP dynamics, while the estimation error of the HAR model parameter is smaller than that of the LASSO-based parameter. That is, the sparsity structure on the HAR model parameters may explain the MVP dynamics well. However, the difference is relatively small; thus, the

LASSO procedure can also capture the MVP dynamics. On the other hand, the model using the CLIME estimator to obtain the inverse matrix performs better than one using the direct inverse operation, such as the FIVAR model. This may be because accommodating the sparsity of the inverse matrix helps account for the MVP weights.

Table 2: Average rank of daily 10-minute portfolio risks for the DRMVP-50, DRMVP-HAR, HAR, FIVAR, and CLIME models for Period 1, Period 2, and the whole period. In the parentheses, we report the number of those ranked first among the models.

Period	DRMVP-50	DRMVP-HAR	HAR	FIVAR	CLIME
1	3.192 (20)	2.800 (11)	2.520 (45)	2.856 (23)	3.632 (26)
2	3.238 (15)	2.524 (17)	2.167 (45)	3.151 (21)	3.921 (28)
Whole	3.215 (35)	2.661 (28)	2.343 (90)	3.004 (44)	3.777 (54)

In addition to evaluating annualized and relative risks, we computed daily risk ranks for the DRMVP-50, DRMVP-HAR, HAR, FIVAR, and CLIME models. Since the rank measure effectively mitigates the undue effect of outlying risks, which can distort annualized and relative risk measures, the rank measure serves as a metric to evaluate whether the proposed model consistently performs better than the other models. Table 2 reports the average rank and the number of those ranked first of daily 10-minute portfolio risks for the DRMVP-50, DRMVP-HAR, HAR, FIVAR, and CLIME models. From Table 2, we find that the HAR model consistently shows the best performance. Meanwhile, the FIVAR model has better average ranks than the DRMVP-50 and CLIME models, whereas the DRMVP-50 and CLIME have smaller annualized risks and mean relative risks. One possible explanation is that the CLIME-based weights are distinct from the FIVAR weights, and the CLIME-based estimators are likely to have similar behavior in terms of the 10-minute portfolio risks. In this situation, the average rank can be more advantageous for FIVAR than DRMVP-50 and CLIME that show consistently worse performance among the CLIME-based estimators.

Table 3 reports the mean ℓ_2-norm of the difference between the ex-post normalized MVP obtained from the CLIME estimator and the estimated normalized MVP using the DRMVP-50, DRMVP-HAR, HAR, FIVAR, and CLIME models for Period 1, Period 2,

Table 3: Mean ℓ_2-norm of the difference between the ex-post normalized MVP obtained from the CLIME estimator and the estimated normalized MVP using the DRMVP-50, DRMVP-HAR, HAR, FIVAR, and CLIME models for Period 1, Period 2, and the whole period.

Period	DRMVP-50	DRMVP-HAR	HAR	FIVAR	CLIME
1	0.154	0.154	0.148	0.169	0.191
2	0.134	0.133	0.127	0.140	0.166
Whole	0.144	0.143	0.138	0.154	0.179

and the whole period. From Table 3, we find that the DR-MVP-based models perform better than the other benchmarks. The HAR model shows the best performance among the DR-MVP-based models. The mean of ℓ_2-norm of the CLIME model is greater than that of the FIVAR model, although the annualized risks and the mean relative risks of the CLIME are less than that of the FIVAR. This may be due to the fact that the FIVAR model captures the linear dynamics of both factor and idiosyncratic volatilities using past volatility matrices. Therefore, by Taylor's expansion, the FIVAR model is able to partially approximate the dynamics of the MVP. From these results, we can conjecture that the proposed DR-MVP model can capture the MVP dynamics.

To compare the predictive accuracy among the DRMVP-50, DRMVP-HAR, HAR, FIVAR, and CLIME models, we conducted Diebold-Mariano (DM) tests (Diebold and Mariano, 2002) for the 10-minute portfolio risks, relative risks, and ℓ_2-norms. We compared the DRMVP-HAR model with other models as follows:

$$H_0 : \mathbb{E}\left[R_i^*\right] = \mathbb{E}\left[R_i\right] \quad \text{vs.} \quad H_1 : \mathbb{E}\left[R_i^*\right] < \mathbb{E}\left[R_i\right],$$

where R_i^* and R_i are one of the 10-minute portfolio risks, relative risks, or ℓ_2-norms for the DRMVP-HAR model and the other models, respectively. Table 4 reports the p-values of the DM tests based on 10-minute portfolio risks, relative risks, and ℓ_2-norm for the DRMVP-HAR model versus one of the DRMVP-50, HAR, FIVAR, and CLIME models

for Period 1, Period 2, and the whole period. From Table 4, we can find that the p-values of DM tests based on 10-minute portfolio risks, relative risks, and ℓ_2-norm are less than 0.1 for all periods, except for the HAR model. From this result, we can conclude that our proposed DR-MVP model performs significantly better than the other models.

Table 4: The p-values of the DM tests based on 10-minute portfolio risks, relative risks, and ℓ_2-norm for the DRMVP-HAR model versus one of the DRMVP-50, HAR, FIVAR, and CLIME models for Period 1, Period 2, and the whole period.

	Risk				Relative risk			
Period	DRMVP-50	HAR	FIVAR	CLIME	DRMVP-50	HAR	FIVAR	CLIME
1	0.086	1.000	0.000	0.014	0.075	1.000	0.000	0.007
2	0.000	1.000	0.000	0.004	0.000	1.000	0.000	0.000
Whole	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000

	ℓ_2-norm			
Period	DRMVP-50	HAR	FIVAR	CLIME
1	0.086	1.000	0.000	0.014
2	0.000	1.000	0.000	0.004
Whole	0.000	1.000	0.000	0.000

To check the economic benefit of predicting MVPs, we calculated the Sharpe ratio as $\bar{ER}/S$, where $\bar{ER}$ and S are the mean and standard deviation of the excess return ER_i, respectively, $ER_i = \widehat{\mathbf{w}}_i^\top R_i - r_{f,i}$, R_i are close-to-close log-return vectors of p-stocks for day i, and $r_{f,i}$ is 3-month T-bill rate for day i. Table 5 reports the out-of-sample Sharpe ratios of the normalized MVP with the DRMVP-50, DRMVP-HAR, HAR, FIVAR, and CLIME models for Period 1, Period 2, and the whole period. Table 5 shows that the proposed DRMVP model does not significantly outperform the other models. This may be because the primary objective of the DR-MVP model is to forecast MVP, which does not necessarily correspond to the point that maximizes the Sharpe ratio on the efficient frontier. It would be an interesting future study to investigate and develop a model to capture the dynamics of the tangency portfolio that maximizes the Sharpe ratio. We leave this for a future study.

Table 5: Out-of-sample Sharpe ratios of the normalized MVP with the DRMVP-50, DRMVP-HAR, HAR, FIVAR, and CLIME models for Period 1, Period 2, and the whole period.

Period	DRMVP-50	DRMVP-HAR	HAR	FIVAR	CLIME
1	0.149	0.155	0.213	0.177	0.231
2	0.152	0.142	0.146	0.086	0.182
Whole	0.151	0.148	0.179	0.133	0.206

6 Conclusion

In this paper, we proposed a novel dynamic realized MVP model that can accommodate nonlinear volatility dynamics by directly explaining the MVP dynamics. To obtain non-parametric realized MVPs for each day, we employed the CLIME estimation procedure under the sparse condition for the inverse volatility matrix. With the non-parametric realized MVP, we found the AR dynamic structure for the realized MVP and constructed dynamic models to explain the AR structure. To connect the high-frequency-based information and low-frequency dynamics, we suggested a diffusion process that would provide the rigorous mathematical background. The empirical study shows that modeling the dynamic structure of the realized MVP helps account for the market dynamics in terms of the minimum variance portfolio.

Acknowledgments

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105(492):1504–1517.

Aït-Sahalia, Y. and Xiu, D. (2016). Increased correlation among asset classes: Are volatility or jumps to blame, or both? *Journal of Econometrics*, 194(2):205–219.

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics*, 89(4):701–720.

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

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

Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics*, 4(1):1–30.

Bibinger, M., Hautsch, N., Malec, P., Reiß, M., et al. (2014). Estimating the quadratic

covariation matrix from noisy observations: Local method of moments and efficiency. *The Annals of Statistics*, 42(4):1312–1346.

Cai, 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(494):594–607.

Cai, T. T., Hu, J., Li, Y., and Zheng, X. (2020). High-dimensional minimum variance portfolio estimation based on high-frequency data. *Journal of Econometrics*, 214(2):482–494.

Chan, L. K., Karceski, J., and Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *The Review of Financial Studies*, 12(5):937–974.

Chanda, K. C. (1974). Strong mixing properties of linear stochastic processes. *Journal of Applied Probability*, 11(2):401–408.

Chen, J. and Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771.

Chen, J. and Chen, Z. (2012). Extended bic for small-n-large-p sparse glm. *Statistica Sinica*, pages 555–574.

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

Clarke, R. G., De Silva, H., and Thorley, S. (2006). Minimum-variance portfolios in the US equity market. *The Journal of Portfolio Management*, 33(1):10–24.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Pirino, D., and Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276–288.

Dai, C., Lu, K., and Xiu, D. (2019). Knowing factors or factor loadings, or neither? evaluating estimators of large covariance matrices with noisy and asynchronous data. *Journal of Econometrics*, 208(1):43–79.

Davies, R. and Tauchen, G. (2018). Data-driven jump detection thresholds for application in jump regressions. *Econometrics*, 6(2):16.

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(5):798–812.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric theory*, 11(1):122–150.

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

Fan, J. and Kim, D. (2018). Robust high-dimensional volatility matrix estimation for high-frequency factor model. *Journal of the American Statistical Association*, 113(523):1268–1283.

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

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.

Fan, J., Liu, H., Sun, Q., and Zhang, T. (2018). I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *Annals of Statistics*, 46(2):814.

Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, 102(480):1349–1362.

Haugen, R. A. and Baker, N. L. (1991). The efficient market inefficiency of capitalization–weighted stock portfolios. *The Journal of Portfolio Management*, 17(3):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(7):2249–2276.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

Kim, D. and Fan, J. (2019). Factor GARCH-Itô models for high-frequency data with application to large volatility matrix prediction. *Journal of Econometrics*, 208(2):395–417.

Kim, D., Kong, X.-B., Li, C.-X., and Wang, Y. (2018a). Adaptive thresholding for large volatility matrix estimation based on high-frequency financial data. *Journal of Econometrics*, 203(1):69–79.

Kim, D., Liu, Y., and Wang, Y. (2018b). Large volatility matrix estimation with factor-based diffusion model for high-frequency financial data. *Bernoulli*, 24(4B):3657–3682.

Kim, D. and Wang, Y. (2016). Sparse PCA based on high-dimensional Itô processes with measurement errors. *Journal of Multivariate Analysis*, 152:172–18.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*, 30(12):4349–4388.

Lee, S. S. and Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies*, 21(6):2535–2563.

Mancini, C. (2004). Estimation of the characteristics of the jumps of a general poisson-diffusion model. *Scandinavian Actuarial Journal*, 2004(1):42–52.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Masini, R. P., Medeiros, M. C., and Mendes, E. F. (2022). Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations. *Journal of Time Series Analysis*, 43(4):532–557.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Pelger, M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics*, 208(1):23–42.

Shin, M., Kim, D., and Fan, J. (2023). Adaptive robust large volatility matrix estimation based on high-frequency financial data. *Journal of Econometrics*, 237(1):105514.

Shin, M., Kim, D., Wang, Y., and Fan, J. (2021). Factor and idiosyncratic VAR-Itô volatility models for heavy-tailed high-frequency financial data. *Available at SSRN 3921526*.

Sun, Q., Zhou, W.-X., and Fan, J. (2020). Adaptive huber regression. *Journal of the American Statistical Association*, 115(529):254–265.

Tao, M., Wang, Y., Zhou, H. H., et al. (2013). Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. *The Annals of Statistics*, 41(4):1816–1864.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of 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. *Journal of Econometrics*, 160(1):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(472):1394–1411.

Zhang, X., Kim, D., and Wang, Y. (2016). Jump variation estimation with noisy high frequency financial data via wavelets. *Econometrics*, 4(3):34.