Shrinkage Estimation of Covariance Matrix for Portfolio Choice with High Frequency Data

Cheng Liu, Ningning Xia and Jun Yu *

November 18, 2016

Abstract

This paper examines the usefulness of high frequency data in estimating the covariance matrix for portfolio choice when the portfolio size is large. A computationally convenient nonlinear shrinkage estimator for the integrated covariance (ICV) matrix of financial assets is developed in two steps. The eigenvectors of the ICV are first constructed from a designed time variation adjusted realized covariance matrix of noise-free log-returns of relatively low frequency data. Then the regularized eigenvalues of the ICV are estimated by quasi-maximum likelihood based on high frequency data. The estimator is always positive definite and its inverse is the estimator of the inverse of ICV. It minimizes the limit of the out-of-sample variance of portfolio returns within the class of rotation-equivalent estimators. It works when the number of underlying assets is larger than the number of time series observations in each asset and when the asset price follows a general stochastic process. Our theoretical results are derived under the assumption that the number of assets (p) and the sample size (n) satisfy $p/n \to y > 0$ as $n \to \infty$. The advantages of our proposed estimator are demonstrated using real data.

Some key words: Portfolio Choice, High Frequency Data; Integrated Covariance Matrix; Shrinkage Function.

JEL classification: C13; C22; C51; G12; G14

*Liu is an assistant professor in Economics and Management School of Wuhan University, Hubei, China. Email: chengliu_eco@whu.edu.cn. Xia is an assistant professor in School of Statistics and Management, Shanghai University of Finance and Economics. Email: xia.ningning@mail.shufe.edu.cn. Yu is a professor in School of Economics and Lee Kong Chian School of Business, Singapore Management University. Email: yujun@smu.edu.sg.

1 Introduction

The portfolio choice problem has been an important topic in modern financial economics ever since the pioneer contribution by Markowitz (1952). It is well-known in the literature that constructing an optimal portfolio requires a good estimate for the second moment of the future return distribution, i.e., the covariance matrix of the future returns. The simplest situation for estimating the covariance matrix is when the returns are independent and identically normally distributed (IID) over time. In this case, the maximum likelihood estimator (MLE) is the sample covariance matrix and the efficiency of MLE is justified asymptotically.

However, there are at least two problems for using the sample covariance matrix to select the optimal portfolio in practice. First, when the portfolio size is large, the sample covariance matrix is found to lead to poor performances in the selected portfolio; see Jobson and Korkie (1980) and Michaud (1989). Not surprisingly, the sample covariance matrix is rarely used by practitioners when the portfolio size is large. The reason for the poor performances is due to the degree-of-freedom argument. That is, too many parameters have to be estimated in the covariance matrix when the portfolio size is large. In fact, if the portfolio size is larger than the number of time series observations in each asset, the sample covariance is always singular. Second, the returns are not IID over time. This is because typically the covariance is time varying. In this case, the asymptotic justification for using the sample covariance matrix is lost.

Many alternative estimators of the large dimensional covariance matrix for portfolio choice have been proposed in the literature. A rather incomplete list includes Ledoit and Wolf (2003, 2004, 2014), Frahm and Memmel (2010), DeMiguel, Garlappi, and Uppal (2009), DeMiguel, Garlappi, Nogales, and Uppal (2009), Kan and Zhou (2007), Fan, Fan and Lv (2008), Pesaran and Zaffaroni (2009), Tu and Zhou (2011). Most studies use dimension reduction techniques. One of the techniques uses factor (either observed factors or latent factors) models. Another approach uses a statistical technique known as shrinkage, a method first introduced by Stein (1956). Murihead (1987) reviewed the literature on shrinkage estimators of the covariance matrix. All these estimators are constructed from low frequency data (daily, weekly or monthly data) over a long period (one year or more). However, if the investment period of a portfolio is much shorter (say one day or one week or one month) which is empirically more relevant, given the time varying nature of the covariance, we expect the covariance in the near future to be similar to the average covariance over an immediate recent time period but not to that over a long time period. Hence, even if data over a long time period is available, one may only prefer using data over a short period. If low frequency data over a short time period are used, however, the degree-of-freedom argument will be applicable.

The recent availability of quality high-frequency data on financial assets has motivated a growing literature devoted to the model-free measurement of covariances. In a recent study, Fan, Li and Yu (2012) proposed to use high-frequency data to estimate the ICV over a short time period for the purpose of portfolio choice. Their setup allows one to impose gross exposure constraints. The use of gross exposure constraints plays a similar role to the no-short-sale constraint in Jagannathan and Ma (2003). Fan, Li and Yu (2012) demonstrated the substantial advantages of using high-frequency date in both simulation and empirical studies.

There are several reasons why it is better to use high frequency data to estimate the covariance matrix. First, the use of high frequency data drastically increases the sample size. This is especially true for liquid assets. Second, one does not need to assume returns are IID any more for establishing the large sample theory for the estimator. This generalization is important due to the time-varying nature of spot covariance. Not surprisingly, the literature on estimating the ICV based on high frequency data is growing rapidly.

In this paper, we also use high frequency data to estimate the ICV for the purpose of portfolio choice. Unlike Fan, Li and Yu (2012) where portfolio choice is done under pre-specified exposure constraints, we focus our attention on how to get a good shrinkage estimator of the ICV without any pre-specified constraint.[1] This shift of focus is due to the lack of guidance on how to specify the gross exposure constraints. Our estimator designs the shrinkage function as in Ledoit and Wolf (2014). However, we differ from Ledoit and Wolf (2014) in the following important ways. First, instead of applying the shrinkage function to the eigenvalues of sample covariance matrix by assuming the returns are IID, we regularize the eigenvalues of a designed time variation adjusted (TVA) realized covariance matrix under the assumption that the covariance matrix is time varying. Second, instead of using low frequency data, we use high frequency data for constructing the designed TVA realized covariance matrix and estimating its regularized eigenvalues. We show that our proposed estimator, which will be given in Section 3, not only has some desirable properties in terms of estimating the ICV, but also asymptotically achieves the minimum out-of-sample portfolio risk.

The paper is organized as follows. In Section 2 we set up the portfolio choice problem. Section 3 introduces our estimator and discusses its properties and implementations. In Section 4, we compare the out-of-sample performance of our proposed method with several methods proposed in the literature using actual data, including the equal weight, the linear shrinkage estimator of Ledoit and Wolf (2004), and the high frequency method of Fan, Li and Yu (2012). Section 5 concludes. The appendix collects the proof of our theoretical results.

2 Portfolio Selection: The Setup

Suppose that a portfolio is constructed based on a pool of p assets whose log-price is denoted by $\mathbf{X}_t = (X_{1t}, \cdots, X_{pt})'$, where $\mathbf{M}'$ denotes the transpose of the vector or matrix $\mathbf{M}$. Instead of assuming $\mathbf{X}_t$ follows a Brownian motion which means that the log-returns are IID, we assume $\mathbf{X}_t$ follows a more general diffusion process as

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Theta}_t d\mathbf{B}_t, \tag{1}$$

[1] DeMiguel, Garlappi, Nogales, and Uppal (2009) showed that adding a constraint for 1-norm of weights is equivalent to shrinkage the estimator of covariance matrix.

where $\boldsymbol{\mu}_t = (\mu_{1t}, \cdots, \mu_{pt})'$ is a p-dimensional drift process at time t, $\boldsymbol{\Theta}_t$ is a $p \times p$ (spot) covolatility matrix at time t, and $\mathbf{B}_t$ is a p-dimensional standard Brownian motion.

A portfolio is constructed based on $\mathbf{X}_t$ with weight $\mathbf{w}_T$ which satisfies $\mathbf{w}_T'\mathbf{1} = 1$ at time T and a holding period τ, where $\mathbf{1}$ is a p-dimensional vector with all elements being 1. Over the period $[T, T+\tau]$, it has a return $\mathbf{w}_T' \int_T^{T+\tau} d\mathbf{X}_t$, and has a risk (variance)

$$R_{T,T+\tau}(\mathbf{w}_T) = \mathbf{w}_T'\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}\mathbf{w}_T, \quad \text{where } \widetilde{\boldsymbol{\Sigma}}_{T,T+\tau} = \int_T^{T+\tau} \mathrm{E}_T \boldsymbol{\Sigma}_t dt,$$

with $\boldsymbol{\Sigma}_t = \boldsymbol{\Theta}_t\boldsymbol{\Theta}_t'$ being the (spot) covariance matrix at time t and E_T denotes the expectation conditional on information up to time T (see Fan, Li and Yu, 2012). Typically, the holding period τ is short (say one day or one week or one month).

To focus on finding a good approximation for $\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}$, we consider the following global minimum variance (GMV) problem:

$$\min_{\mathbf{w}_T} \mathbf{w}_T'\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}\mathbf{w}_T \quad \text{with} \quad \mathbf{w}_T'\mathbf{1} = 1. \tag{2}$$

By taking the derivative of $\mathbf{w}_T$, we have the following theoretical optimal weight,

$$\mathbf{w}_T = \frac{\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}^{-1}\mathbf{1}}{\mathbf{1}'\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}^{-1}\mathbf{1}}, \tag{3}$$

which is a function of the expected ICV conditional on the current time T, i.e., $\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}$.

Denote the ICV over the period $[T-h, T]$ by

$$\boldsymbol{\Sigma}_{T-h,T} := \int_{T-h}^{T} \boldsymbol{\Sigma}_t dt.$$

If h is small, following Fan, Li and Yu (2012), we use the following approximation

$$\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau} \approx \frac{\tau}{h}\, \boldsymbol{\Sigma}_{T-h,T}. \tag{4}$$

Consequently, the theoretical optimal weight becomes

$$\mathbf{w}_T = \frac{\boldsymbol{\Sigma}_{T-h,T}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}_{T-h,T}^{-1}\mathbf{1}}. \tag{5}$$

The reason for choosing a small h from the historical sample (i.e. a small time span for $[T-h, T]$) to approximate the expected ICV is due to the time varying and persistent nature of the covariance matrix. If a big h (say 10 years) is used and an average covariance matrix is used to approximate the expected ICV, the approximation errors would be inevitably large. In fact, as rightly argued in Fan, Li and Yu (2012), even when the true covariance matrices are available, an average of them will still lead to large approximation errors.

Let $\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}$ denote a generic (invertible) estimator of the ICV $\boldsymbol{\Sigma}_{T-h,T}$. The plug-in estimator of the optimal portfolio weight for $\mathbf{w}_T$ in (5) is

$$\hat{\mathbf{w}}^*_T := \frac{\left(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}\right)^{-1}\mathbf{1}}{\mathbf{1}'\left(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}\right)^{-1}\mathbf{1}}.$$

We need to find the optimal $\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}$ for portfolio choice. Given that the optimal portfolio is typically meant to perform the best out-of-sample, following Ledoit and Wolf (2014), we define a loss function for portfolio selection to be the out-of-sample variance of portfolio returns conditional on $\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}$,

$$\mathcal{L}(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}, \boldsymbol{\Sigma}_{T-h,T}) = (\hat{\mathbf{w}}^*_T)'\,\boldsymbol{\Sigma}_{T-h,T}\hat{\mathbf{w}}^*_T = \frac{\mathbf{1}'\left(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\left(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}\right)^{-1}\mathbf{1}}{\left\{\mathbf{1}'\left(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}\right)^{-1}\mathbf{1}\right\}^2}, \qquad (6)$$

where we approximate $\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}$ by $\frac{\tau}{h}\boldsymbol{\Sigma}_{T-h,T}$ and ignore the scale $\frac{\tau}{h}$ without any loss. The best estimator of the ICV is therefore the one that minimizes the loss function $\mathcal{L}(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}, \boldsymbol{\Sigma}_{T-h,T})$.

Although this paper mainly focuses on the GMV problem, our estimation technique has a much wider implications for other problems that also require the estimation of ICV, including the Markowitz portfolios with and without estimating the conditional mean. In the empirical studies, we will show the usefulness of our proposed method in the context of the Markowitz portfolio.

3 The New Estimator of ICV

Denote the trading time points for the ith asset by $0 \le t_{i1} < t_{i2} < ... < t_{i,N_i} \le T$ with $i = 1, ..., p$. It is difficult to estimate the ICV based on tick-by-tick high frequency data when the number of stocks (p) is large for the following reasons. First, data are always non-synchronous. Second, data are contaminated by microstructure noises. Denote $Y_{i,t_{ij}}$ the log-price of the ith asset at time t_{ij} and $X_{i,t_{ij}}$ the latent log efficient price of the ith asset. Then

$$Y_{i,t_{ij}} = X_{i,t_{ij}} + \epsilon_{i,t_{ij}},$$

where $\epsilon_{i,t_{ij}}$ is the market microstructure noise at time t_{ij}. Third, the spot covariance matrix $\boldsymbol{\Sigma}_t$ of returns of latent log-price $\mathbf{X}_t$ is time varying. Fourth, the returns of the efficient price are not independent over time. To find a good estimator for the ICV, we first introduce an initial estimator, denoted the time variation adjust (TVA) realized covariance matrix, and discuss its disadvantages for estimating the ICV in subsection 3.1. To improve the initial estimator, we propose to regularize its eigenvalues. In subsection 3.2, we provide the theoretical background for regularizing the eigenvalues of TVA realized covariance matrix. We then demonstrate how to regularize its eigenvalues in subsection 3.3.

3.1 The initial estimator of ICV: TVA

To simplify the problem, we propose the following structural assumption for $\mathbf{X}_t$. The same assumption was also used in Zheng and Li (2011).

Definition 3.1. *(Class $\mathcal{C}$). Suppose that $\mathbf{X}_t$ is a p-dimensional process satisfying Equation* (1)*. We say that $\mathbf{X}_t$ belongs to class $\mathcal{C}$ if, almost surely, there exist $\gamma_t \in D([T-h,T];\mathbb{R})$ and $\mathbf{\Lambda}$ a $p\times p$ matrix satisfying* $\operatorname{tr}(\mathbf{\Lambda}\mathbf{\Lambda}') = p$ *such that*

$$\mathbf{\Theta}_t = \gamma_t \mathbf{\Lambda},$$

where $D([T-h,T];\mathbb{R})$ stands for the space of càdlàg functions from $[T-h,T]$ to $\mathbb{R}$.

Remark 3.1. *Class $\mathcal{C}$ allows the covariance matrix to be time varying because γ_t is time varying. The assumption of $\mathbf{\Theta}_t = \gamma_t \mathbf{\Lambda}$ may be too strong than necessary but facilitates the mathematical proof of the results in the present paper.*

If $\mathbf{X}_t$ belongs to class $\mathcal{C}$, we can decompose

$$\mathbf{\Sigma}_{T-h,T} = \int_{T-h}^{T} \gamma_t^2 dt \cdot \mathbf{\Lambda}\mathbf{\Lambda}' = \mathbf{P}\left(\int_{T-h}^{T} \gamma_t^2 dt \cdot \mathbf{\Gamma}\right)\mathbf{P}',$$

where $\mathbf{\Gamma}$ is a diagonal matrix, $\mathbf{P}$ an orthogonal matrix, and $\mathbf{P}\mathbf{\Gamma}\mathbf{P}'$ the eigen-decomposition of $\mathbf{\Lambda}\mathbf{\Lambda}'$ such that the eigenvalues and eigenvectors of $\mathbf{\Sigma}_t = \mathbf{\Theta}_t\mathbf{\Theta}_t'$ are time varying and invariant respectively.

To estimate $\mathbf{\Sigma}_{T-h,T}$, Zheng and Li (2011) proposed to use the so-called TVA realized covariance matrix over the period $[T-h,T]$, which is defined as

$$\mathbf{S}_{T-h,T}^{\mathrm{TVA}} = \frac{\operatorname{tr}\left(\sum_{k=1}^{n} \Delta\mathbf{X}_k \Delta\mathbf{X}_k'\right)}{p} \cdot \breve{\mathbf{S}}_{T-h,T}, \quad \text{where } \breve{\mathbf{S}}_{T-h,T} = \frac{p}{n}\sum_{k=1}^{n} \frac{\Delta\mathbf{X}_k \Delta\mathbf{X}_k'}{|\Delta\mathbf{X}_k|^2}, \tag{7}$$

$\Delta\mathbf{X}_k = \mathbf{X}_{\tau_k} - \mathbf{X}_{\tau_{k-1}}$, and $\mathbf{X}_{\tau_k}$ denotes the log efficient price $\mathbf{X}_t$ at time τ_k for

$$T-h := \tau_0 < \tau_1 < \cdots < \tau_n := T.$$

Zheng and Li (2011) demonstrated that $\operatorname{tr}\left(\sum_{k=1}^{n} \Delta\mathbf{X}_k \Delta\mathbf{X}_k'\right)/p$ is a good estimator for $\int_{T-h}^{T} \gamma_t^2 dt$ and $\breve{\mathbf{S}}_{T-h,T}$ is similar to the sample covariance matrix with IID samples. Here similarity means that $\breve{\mathbf{S}}_{T-h,T}$ is a consistent estimator of population covariance matrix $\mathbf{\Lambda}\mathbf{\Lambda}'$ when p is fixed, while the limiting spectral distribution of $\breve{\mathbf{S}}_{T-h,T}$, which will be introduced later in the paper, is equivalent to that of the sample covariance matrix of IID samples generated from a distribution with zeros mean and population covariance $\mathbf{\Lambda}\mathbf{\Lambda}'$, when p goes to ∞ together with the sample size n.

Clearly, the construction of TVA requires a synchronous record of p assets at $(\tau_0, \tau_1, \cdots, \tau_n)$. Since data is always non-synchronous, we need to synchronize them. In this paper, we use the previous tick method (see Zhang, 2011) to interpolate the prices. However, the efficient price is latent due to the presence of microstructure noise. To deal with this problem, we suggest using sparse sampling so that the impact of microstructure noise can be ignored. Based on a Hausman type test, Aït-Sahalia and Xiu (2016) showed that when data are sampled every 15 minutes, the observed prices are free of the microstructure noise problem. In this paper, we will follow this suggestion by sampling the interpolated data every 15 minutes. Denote $(\tau_0, \tau_1, \cdots, \tau_n)$ the time stamps at every 15 minutes. So $\mathbf{Y}_{\tau_k} \approx \mathbf{X}_{\tau_k}$.

Denote the sparsely-sampled log-prices by $\mathbf{Y}_{\tau_0}, \mathbf{Y}_{\tau_1}, ..., \mathbf{Y}_{\tau_n}$. The feasible TVA realized covariance matrix is constructed as

$$\widetilde{\mathbf{S}}_{T-h,T}^{\mathrm{TVA}} = \frac{\operatorname{tr}\left(\sum_{k=1}^{n} \Delta\mathbf{Y}_k \Delta\mathbf{Y}_k'\right)}{n} \sum_{k=1}^{n} \frac{\Delta\mathbf{Y}_k \Delta\mathbf{Y}_k'}{|\Delta\mathbf{Y}_k|^2}, \tag{8}$$

with $\Delta\mathbf{Y}_k = \mathbf{Y}_{\tau_k} - \mathbf{Y}_{\tau_{k-1}}$. Since $\widetilde{\mathbf{S}}_{T-h,T}^{\mathrm{TVA}}$ has the same properties as $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$, we treat $\widetilde{\mathbf{S}}_{T-h,T}^{\mathrm{TVA}}$ the same as $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$ and only use $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$ in the rest of this paper.

It is well-known that the eigenvalues of the sample covariance matrix are more spread out than those of the population covariance matrix. This property is applicable not only to the sample covariance matrix but also to $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$. In other words, the smallest eigenvalues of $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$ tend to be biased downwards, while the largest ones upwards. As a result, there is a need to regularize the eigenvalues of $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$.

3.2 Theoretical background for regularizing the eigenvalues of $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$

Let us first introduce some concepts in the random matrix theory. Let p denote the number of variables and $n = n(p)$ the sample size. For any $p \times p$ symmetric matrix $\mathbf{M}$, suppose that its eigenvalues are $\lambda_1, \cdots, \lambda_p$, sorted in the non-increasing order. Then the empirical spectral distribution (ESD) of $\mathbf{M}$ is defined as

$$F^{\mathbf{M}}(x) := \frac{1}{p} \sum_{i=1}^{p} \mathbb{I}(\lambda_i \leq x), \qquad \text{for} \quad x \in \mathbb{R},$$

where $\mathbb{I}$ denotes the indicator function of a set. The limit of ESD as $p \to \infty$, if exists, is referred to the limiting spectral distribution (LSD hereafter). Let Supp(G) denotes the support interval of distribution function G. For any distribution G, $s_G(\cdot)$ denotes its Stieltjes transform defined as

$$s_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \qquad \text{for } z \in \mathbb{C}^+ := \{z \in \mathbb{C} : \Im(z) > 0\},$$

where $\Im(\cdot)$ denotes the imaginary part of a complex number.

3.2.1 The limit of loss function

Suppose the eigen-decomposition of $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$ is

$$\mathbf{S}_{T-h,T}^{\mathrm{TVA}} = \mathbf{U}\mathbf{V}\mathbf{U}' = \mathbf{U}\operatorname{diag}(v_1, ..., v_p)\mathbf{U}', \tag{9}$$

where $v_1, ..., v_p$ are eigenvalues of $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$ sorted in the non-increasing order, $\mathbf{U} = (\mathbf{u}_1, ..., \mathbf{u}_p)$ are corresponding eigenvectors. Let $\operatorname{diag}(\mathbf{M})$ denote a diagonal matrix with the diagonal elements being the diagonal elements of $\mathbf{M}$ if $\mathbf{M}$ is a matrix or being $\mathbf{M}$ if $\mathbf{M}$ is a vector.

To regularize the eigenvalues of $\mathbf{S}^{\text{TVA}}_{T-h,T}$, following Ledoit and Wolf (2014), we restrict our attention to a class of rotation-equivalent estimators which is defined below. This strategy allows us to use a nonlinear shrinkage method to regularize the eigenvalues. However, different from Ledoit and Wolf (2014), we do not assume returns are IID. Instead we assume that $\mathbf{X}_t \in \mathcal{C}$.

Definition 3.2. *(Class of Estimators $\mathcal{S}$). We consider a generic positive definite estimator for $\boldsymbol{\Sigma}_{T-h,T}$ of the type $\widehat{\boldsymbol{\Sigma}}^*_{T-h,T} := \mathbf{U}\text{diag}(g_n(v_1), \cdots, g_n(v_p))\mathbf{U}'$, with $v_1 \geq \cdots \geq v_p$ being the eigenvalues of $\mathbf{S}^{\text{TVA}}_{T-h,T}$, $\mathbf{U} = (\mathbf{u}_1, ..., \mathbf{u}_p)$ being corresponding eigenvectors. Here g_n is a real univariate function and can depend on $\mathbf{S}^{\text{TVA}}_{T-h,T}$. We assume that there exists a nonrandom real univariate function $g(x)$, defined on* Supp(F) *and continuously differentiable, such that $g_n(x) \xrightarrow{a.s.} g(x)$, for all $x \in$* Supp(F)*, where F denotes the LSD of $\mathbf{S}^{\text{TVA}}_{T-h,T}$.*

Here, $g_n(x)$ is called the *shrinkage function* because what it does is to shrink the eigenvalues of $\mathbf{S}^{\text{TVA}}_{T-h,T}$ by reducing the dispersion around the mean, pushing up the small ones and pulling down the large ones. The high dimensional asymptotic properties of $\mathbf{S}^{\text{TVA}}_{T-h,T}$ are fully characterized by its limiting shrinkage function $g(x)$. As noted in Stein (1975) and Ledoit and Wolf (2014), the estimators in this class are rotation equivalent, a property that is desired when the user does not have any prior preference about the orientation of the eigenvectors.

Since we consider the case that p goes to ∞ together with the sample size, finding the optimal estimator of $\boldsymbol{\Sigma}_{T-h,T}$ within class $\mathcal{S}$ for portfolio selection is equivalent to finding the optimal shrinkage function $g(x)$ that minimizes the limit of the loss function $\mathcal{L}\left(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}, \boldsymbol{\Sigma}_{T-h,T}\right)$ for $\widehat{\boldsymbol{\Sigma}}^*_{T-h,T} \in \mathcal{S}$. We have the following theorem to show the limit of $\mathcal{L}\left(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}, \boldsymbol{\Sigma}_{T-h,T}\right)$.

Theorem 3.1. *Suppose that $\mathbf{X}_t$ is a p -dimensional diffusion process in class $\mathcal{C}$ for some drift process $\boldsymbol{\mu}_t$, covolatility process $\boldsymbol{\Theta}_t = \gamma_t \boldsymbol{\Lambda}$ and p-dimensional Brownian motion $\mathbf{B}_t$, which satisfies the following assumptions:*

(A.i) *$\boldsymbol{\mu}_t = 0$ for $t \in [T-h, T]$, and γ_t is independent of $\mathbf{B}_t$.*

(A.ii) *There exists $C_0 < \infty$ such that for all p, $|\gamma_t| \in (1/C_0, C_0)$ for all $t \in [T-h, T]$ almost surely;*

(A.iii) *All eigenvalues of $\breve{\boldsymbol{\Sigma}} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ are bounded uniformly from 0 and infinity;*

(A.iv) $\lim_{p\to\infty} \text{tr}\left(\boldsymbol{\Sigma}_{T-h,T}\right)/p = \lim_{p\to\infty} \int_{T-h}^{T} \gamma_t^2 dt := \theta > 0$ *almost surely;*

(A.v) *Almost surely, as $p \to \infty$, the ESD of $\boldsymbol{\Sigma}_{T-h,T}$ converges to a probability distribution H on a finite support;*

(A.vi) *The observation time points τ_k's are independent of the Brownian motion $\mathbf{B}_t$ and there exists a constant $C_1 > 0$ such that* $\max_{1\leq k\leq n} n(\tau_k - \tau_{k-1}) \leq C_1$.

If $p/n \to y \in (0, \infty)$, then the ESD of $\mathbf{S}^{\text{TVA}}_{T-h,T}$ converges almost surely to a nonrandom probability distribution F. If Equation (9) is satisfied, then

$$p \times \mathcal{L}\left(\widehat{\boldsymbol{\Sigma}}^*_{T-h,T}, \boldsymbol{\Sigma}_{T-h,T}\right) \xrightarrow{a.s.} \int \frac{x}{|1-y-yx\times \breve{s}_F(x)|^2 g(x)} dF(x) / \left(\int \frac{dF(x)}{g(x)}\right)^2,$$

where $\widehat{\mathbf{\Sigma}}^*_{T-h,T} := \mathbf{U}\mathrm{diag}(g_n(v_1), \cdots, g_n(v_p))\mathbf{U}'$ *is in class* $\mathcal{S}$ *by regularizing* $\mathbf{S}^{\mathrm{TVA}}_{T-h,T}$, $g(x)$ *is the limiting shrinkage function of* $\widehat{\mathbf{\Sigma}}^*_{T-h,T}$. *In addition, for all* $x \in (0, \infty)$, $\breve{s}_F(x)$ *is defined as* $\lim_{z \in \mathbb{C}^+ \to x} s_F(z)$, *and* $s_F(z)$ *is the Stieltjes transform of the limiting spectral distribution of* $\mathbf{S}^{\mathrm{TVA}}_{T-h,T}$.

Remark 3.2. *Theorem 3.1 extends the result in Proposition 3.1 of Ledoit and Wolf (2014) from the IID case to Class* $\mathcal{C}$ *and from the sample covariance to the TVA realized covariance.*

Remark 3.3. *Without loss of generality, if we assume that all the eigenvalues of* $\widehat{\mathbf{\Sigma}}^*_{T-h,T}$ *and* $\mathbf{\Sigma}_{T-h,T}$ *are bounded,* $\mathbf{1}'\left(\widehat{\mathbf{\Sigma}}^*_{T-h,T}\right)^{-1}\mathbf{\Sigma}_{T-h,T}\left(\widehat{\mathbf{\Sigma}}^*_{T-h,T}\right)^{-1}\mathbf{1} = O_p(p)$ *and* $\mathbf{1}'\left(\widehat{\mathbf{\Sigma}}^*_{T-h,T}\right)^{-1}\mathbf{1} = O_p(p)$, *so that* $\mathcal{L}\left(\widehat{\mathbf{\Sigma}}^*_{T-h,T}, \mathbf{\Sigma}_{T-h,T}\right) = O_p(\frac{1}{p})$. *This is why we investigate the limiting behavior of* $p \times \mathcal{L}\left(\widehat{\mathbf{\Sigma}}^*_{T-h,T}, \mathbf{\Sigma}_{T-h,T}\right)$ *in Theorem 3.1.*

Lemma 3.1. *Under the assumptions of Theorem 3.1, a generic positive-definite estimator* $\widehat{\mathbf{\Sigma}}^*_{T-h,T}$ *within class* $\mathcal{S}$ *minimizes the almost sure limit of the loss function* $\mathcal{L}\left(\widehat{\mathbf{\Sigma}}^*_{T-h,T}, \mathbf{\Sigma}_{T-h,T}\right)$ *if and only if its limiting shrinkage function* g *satisfies*

$$g(x) = \frac{x}{|1 - y - yx \times \breve{m}_F(x)|^2}, \qquad \forall\ x \in \mathrm{Supp(F)}. \tag{10}$$

Lemma 3.1 is a direct conclusion from Theorem 3.1 and Proposition 4.1 of Ledoit and Wolf (2014). Unfortunately, the above minimization problem does not yield a closed-form solution for $g(x)$ because of $\breve{m}_F(x)$ is unknown. In addition, finding $\breve{m}_F(x)$ and then $g(x)$ is numerically difficult in practice. Finding a good algorithm for estimating $\breve{m}_F(x)$ is of great interest as it was done in Ledoit and Wolf (2014) that used a commercial package. However, in this paper we propose to find an alternative interpretation of $g(x)$, which offers an easier way to approximate $g(x)$.

3.2.2 Alternative interpretation of $g(x)$

Motivated from Ledoit and Pèchè (2011), we can show that $g(x)$ in (10) is equivalent to the asymptotic quantity corresponding to the oracle nonlinear shrinkage estimator derived from the following Frobenius norm of the difference between $\mathbf{U}\widetilde{\mathbf{V}}\mathbf{U}'$ and $\mathbf{\Sigma}_{T-h,T}$, i.e.,

$$\min_{\widetilde{\mathbf{V}}\ \text{diagonal}} \|\mathbf{U}\widetilde{\mathbf{V}}\mathbf{U}' - \mathbf{\Sigma}_{T-h,T}\|_F,$$

where the Frobenius norm is defined as $\|\mathbf{M}\|_F = \sqrt{\mathrm{tr}(\mathbf{M}\mathbf{M}')}$ for any real matrix $\mathbf{M}$.

Elementary matrix algebra shows that the solution is

$$\widetilde{\mathbf{V}} = \mathrm{diag}(\tilde{v}_1, \cdots, \tilde{v}_p), \qquad \text{where } \tilde{v}_i = \mathbf{u}_i'\mathbf{\Sigma}_{T-h,T}\mathbf{u}_i,\ i = 1, \cdots, p. \tag{11}$$

To characterize the asymptotic behavior of $\tilde{v}_i, i = 1, \cdots, p$, following the idea of Ledoit and Pèchè (2011), we define the following non-decreasing function

$$\Psi_p(x) = \frac{1}{p}\sum_{i=1}^{p} \tilde{v}_i\ \mathbb{I}(v_i \le x) = \frac{1}{p}\sum_{i=1}^{p} \mathbf{u}_i'\mathbf{\Sigma}_{T-h,T}\mathbf{u}_i \cdot \mathbb{I}(v_i \le x). \tag{12}$$

Theorem 3.2. *Assume that assumptions (A.i)-(A.vi) in Theorem 3.1 hold true and let Ψ_p be defined as in* (12)*. If $p/n \to y \in (0, \infty)$, then there exists a nonrandom function Ψ defined over $\mathbb{R}$ such that $\Psi_p(x)$ converges almost surely to $\Psi(x)$ for all $x \in \mathbb{R} \backslash \{0\}$. If in addition $y \neq 1$, then Ψ can be expressed as*

$$\forall\ x \in \mathbb{R}, \quad \Psi(x) = \int_{-\infty}^{x} \delta(v) dF(v), \tag{13}$$

where F is the LSD of $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$, and if $v > 0$,

$$\delta(v) = \frac{v}{|1 - y - yv \times \breve{m}_F(v)|^2}.$$

Remark 3.4. *Theorem 3.2 extends the result in Theorem 4 of Ledoit and Pèchè (2011) from the IID case to Class $\mathcal{C}$.*

Theorem 3.2 implies that the asymptotic quantity that corresponds to $\tilde{v}_i = \mathbf{u}_i' \boldsymbol{\Sigma}_{T-h,T} \mathbf{u}_i$ is $\delta(v)$ provided that v corresponds to v_i. An interesting finding is that the results of Lemma 3.1 and Theorem 3.2 are consistent with each other, even though they are motivated from two different perspectives. Given that it is much easier to work on the minimization problem in (11), we recommend to regularize the eigenvalues of $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$ by using (11), which is to find a good estimator for each $\tilde{v}_i = \mathbf{u}_i' \boldsymbol{\Sigma}_{T-h,T} \mathbf{u}_i$ with $i = 1, ..., p$.

3.3 Regularized estimators of eigenvalues of $\mathbf{S}_{T-h,T}^{\mathrm{TVA}}$

Note that $\tilde{v}_i = \mathbf{u}_i' \boldsymbol{\Sigma}_{T-h,T} \mathbf{u}_i$ is actually the integrated volatility of process $\mathbf{u}_i' \mathbf{X}_t$ over $[T-h, T]$ for $i = 1, 2, \cdots, p$. A natural estimator of each $\tilde{v}_i$ is the realized volatility $\sum_{k=1}^{n} (\mathbf{u}_i' \Delta \mathbf{X}_k)^2$. Unfortunately, this is not a good idea. To see the problem, note that

$$\widehat{\boldsymbol{\Sigma}}_{T-h,T}^{**} = \mathbf{U} \mathrm{diag} \left(\sum_{k=1}^{n} (\mathbf{u}_1' \Delta \mathbf{X}_k)^2, ..., \sum_{k=1}^{n} (\mathbf{u}_p' \Delta \mathbf{X}_k)^2 \right) \mathbf{U}'.$$

Let us consider the simplest case where $\gamma_t = 1$, $\boldsymbol{\Lambda} = \mathbf{I}_p$ with $\mathbf{I}_p$ be a p-dimensional identity matrix, and $\tau_k - \tau_{k-1} = \frac{h}{n}$ for $k = 1, ..., n$. We can write $\Delta \mathbf{X}_k = \left(\frac{h}{n}\right)^{1/2} \mathbf{Z}_k$ with $\mathbf{Z}_k$'s are IID p-dimensional standard normals such that $\dfrac{\Delta \mathbf{X}_k \Delta \mathbf{X}_k'}{|\Delta \mathbf{X}_k|^2} = \frac{\mathbf{Z}_k \mathbf{Z}_k'}{|\mathbf{Z}_k|^2}$. Since $|\mathbf{Z}_k|^2 \sim p$ as $p \to \infty$, we have

$$\begin{aligned} \mathbf{S}_{T-h,T}^{\mathrm{TVA}} &= \frac{\mathrm{tr}\left(\sum_{k=1}^{n} \Delta \mathbf{X}_k \Delta \mathbf{X}_k' \right)}{p} \frac{p}{n} \sum_{k=1}^{n} \frac{\Delta \mathbf{X}_k \Delta \mathbf{X}_k'}{|\Delta \mathbf{X}_k|^2} \\ &\sim \frac{\mathrm{tr}\left(\sum_{k=1}^{n} \Delta \mathbf{X}_k \Delta \mathbf{X}_k' \right)}{p} \frac{1}{n} \sum_{k=1}^{n} \mathbf{Z}_k \mathbf{Z}_k', \end{aligned}$$

$$\sum_{k=1}^{n} \Delta \mathbf{X}_k \Delta \mathbf{X}_k' = \frac{h}{n} \sum_{k=1}^{n} \mathbf{Z}_k \mathbf{Z}_k'.$$

By denoting $\Delta\mathbf{X} = (\Delta\mathbf{X}_1, ..., \Delta\mathbf{X}_n)'$, we have

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}^{**}_{T-h,T} &= \mathbf{U}\text{diag}\left(\sum_{k=1}^{n}(\mathbf{u}_1'\Delta\mathbf{X}_k)^2, ..., \sum_{k=1}^{n}(\mathbf{u}_p'\Delta\mathbf{X}_k)^2\right)\mathbf{U}' \\
&= \mathbf{U}\text{diag}\left(\mathbf{u}_1'\Delta\mathbf{X}\Delta\mathbf{X}'\mathbf{u}_1, ..., \mathbf{u}_p'\Delta\mathbf{X}\Delta\mathbf{X}'\mathbf{u}_p\right)\mathbf{U}' \\
&= \mathbf{U}\text{diag}\left(\mathbf{U}'\Delta\mathbf{X}\Delta\mathbf{X}'\mathbf{U}\right)\mathbf{U}' \\
&\sim \Delta\mathbf{X}\Delta\mathbf{X}',
\end{aligned}$$

which is actually the sample covariance matrix of IID samples generated from $N(\mathbf{0}, h\mathbf{I}_p)$. Hence, its eigenvalues are also more spread out than that of $h\mathbf{I}_p$, a well-known result in the literature.

To solve this problem, we use the idea from Abadir et al. (2014) and Lam (2016) by splitting the sample into two parts. We use the estimated eigenvectors from a fraction of the data to transform the data into approximately orthogonal series.[2] We then use the independence of two sample covariance matrices to regularize the eigenvalues of one of them. Therefore, instead of based $\mathbf{U}$ on $\Delta\mathbf{X}_k = \mathbf{X}_{\tau_k} - \mathbf{X}_{\tau_{k-1}}$ $(k = 1, ..., n)$ for $T - h := \tau_0 < ... < \tau_n := T$, we base $\mathbf{U}^*$ on $\Delta\mathbf{X}^*_r = \mathbf{X}_{\tau^*_r} - \mathbf{X}_{\tau^*_{r-1}}(r = 1, ..., m)$ for

$$0 := \tau^*_0 < \tau^*_1 < ... < \tau^*_m < T - h,$$

where $\mathbf{U}^* = (\mathbf{u}^*_1, ..., \mathbf{u}^*_p)$ are the eigenvectors of $\mathbf{S}^{\text{TVA}}_{0,T-h}$ corresponding to the eigenvalues with the non-increasing order, and the TVA realized covariance matrix

$$\mathbf{S}^{\text{TVA}}_{0,T-h} = \frac{\text{tr}\left\{\sum_{r=1}^{m}\Delta\mathbf{X}^*_r\left(\Delta\mathbf{X}^*_r\right)'\right\}}{p}\cdot\breve{\mathbf{S}}_{0,T-h}, \quad \text{with} \quad \breve{\mathbf{S}}_{0,T-h} = \frac{p}{m}\sum_{r=1}^{m}\frac{\Delta\mathbf{X}^*_r\left(\Delta\mathbf{X}^*_r\right)'}{|\Delta\mathbf{X}^*_r|^2}.$$

In addition, since the eigenvectors of $\boldsymbol{\Sigma}_t$ is assumed to be time invariant, we also consider the following optimization problem

$$\min_{\mathbf{V}^* \text{ diagonal}} \|\mathbf{U}^*\mathbf{V}^*\left(\mathbf{U}^*\right)' - \boldsymbol{\Sigma}_{T-h,T}\|_F,$$

and estimate each diagonal element of the oracle minimizer $\mathbf{V}^* = \text{diag}(v^*_1, ..., v^*_p)$ with $v^*_i = \left(\mathbf{u}^*_i\right)'\boldsymbol{\Sigma}_{T-h,T}\mathbf{u}^*_i$ based on the data over the time period $[T-h, T]$. To get an accurate estimator for each v^*_i with $i \in \{1, ..., p\}$, we propose to use all the tick-by-tick high frequency data and take into account with the microstructure noises.

Let us first consider the case that the data are synchronous and equally recorded at time points $\{T - h := t^*_0 < t^*_1 < \cdots < t^*_N := T\}$, where the time interval $\Delta = t^*_j - t^*_{j-1} \to 0$ for all $j = 1, ..., N$ as $N \to \infty$ and h fixed. Notice that here $\{t^*_j : j = 0, ..., N\}$ may be quite different from $\{\tau_k : k = 0, ..., n\}$ and Δ can be one second or a few seconds, and should be much smaller than $\tau_k - \tau_{k-1}$ which is 15 minutes.

[2] Strictly speaking, the asymptotic justification of the method requires the IID assumption as shown in Lam (2016). While the IID assumption does not hold for Class $\mathcal{C}$, we examine the effectiveness of this method using real data later.

We assume each observation is contaminated by microstructure noise such that $\mathbf{Y}_t = (Y_{1t}, ..., Y_{pt})'$ (observed) contains the true log-price $\mathbf{X}_t$ (latent) and the microstructure noise $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{pt})'$ in an additive form

$$\mathbf{Y}_t = \mathbf{X}_t + \boldsymbol{\epsilon}_t, \text{ for } t \in [T-h, T], \tag{14}$$

where the p-dimensional noise $\boldsymbol{\epsilon}_t$ is assumed to satisfy

Assumption 1. *The p-dimensional noise $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{pt})'$ at different time points $t = t_0^*, t_1^*, \cdots, t_N^*$ are IID random vectors with mean $\mathbf{0}$ (a p-dimensional vector with all elements being 0), positive definite covariance matrix $\mathbf{A}_0$ and finite fourth moment. In addition, $\boldsymbol{\epsilon}_t$ and $\mathbf{X}_t$ are mutually independent.*

This assumption has commonly been used in the literature; see, for example, Aït-Sahalia et al. (2010), Zhang (2011), Liu and Tang (2014). To estimate $(\mathbf{u}_i^*)'\boldsymbol{\Sigma}_{T-h,T}\mathbf{u}_i^*$, we apply the quasi-maximum likelihood (QML) approach developed in Xiu (2010). Based on (1) and (14), we have

$$\begin{aligned} \tilde{Y}_{it} &= (\mathbf{u}_i^*)'\mathbf{Y}_t = (\mathbf{u}_i^*)'\mathbf{X}_t + (\mathbf{u}_i^*)'\boldsymbol{\epsilon}_t = \tilde{X}_{it} + \tilde{\epsilon}_{it} \\ d\tilde{X}_{it} &= (\mathbf{u}_i^*)'d\mathbf{X}_t = (\mathbf{u}_i^*)'\boldsymbol{\mu}_t dt + (\mathbf{u}_i^*)'\boldsymbol{\Theta}_t d\mathbf{B}_t = \tilde{\mu}_{it}dt + \tilde{\sigma}_{it}d\tilde{B}_{it} \end{aligned} \tag{15}$$

by letting

$$\begin{aligned} \tilde{X}_{it} &= (\mathbf{u}_i^*)'\mathbf{X}_t, \qquad \tilde{\epsilon}_{it} = (\mathbf{u}_i^*)'\boldsymbol{\epsilon}_t, \qquad \tilde{\mu}_{it} = (\mathbf{u}_i^*)'\boldsymbol{\mu}_t, \\ \tilde{\sigma}_{it}d\tilde{B}_{it} &= (\mathbf{u}_i^*)'\boldsymbol{\Theta}_t d\mathbf{B}_t, \quad \tilde{\sigma}_{it}^2 = (\mathbf{u}_i^*)'\boldsymbol{\Theta}_t((\mathbf{u}_i^*)'\boldsymbol{\Theta}_t)' = (\mathbf{u}_i^*)'\boldsymbol{\Theta}_t\boldsymbol{\Theta}_t'\mathbf{u}_i^* = (\mathbf{u}_i^*)'\boldsymbol{\Sigma}_t\mathbf{u}_i^*, \end{aligned}$$

such that $v_i^* = \int_{T-h}^T \tilde{\sigma}_{it}^2 dt$.

Ignoring the impact of $\tilde{\mu}_{it}dt$ by considering $\tilde{\mu}_{it} = 0$, we follow the idea in Xiu (2010) to give two misspecified assumptions for each $i \in \{1, ..., p\}$. First, the spot volatility is assumed to be time invariant: $\tilde{\sigma}_{it}^2 = (\mathbf{u}_i^*)'\boldsymbol{\Sigma}_t\mathbf{u}_i^* = \tilde{\sigma}_i^2$. Second, the noise $\tilde{\epsilon}_{it}$ is assumed to be normally distributed with mean 0 and variance $\tilde{a}_i^2$. Then the quasi-log likelihood function for $\tilde{Y}_{i,t_j^*} - \tilde{Y}_{i,t_{j-1}^*}$ is

$$l(\tilde{\sigma}_i^2, \tilde{a}_i^2) = -\frac{1}{2}\log\det(\boldsymbol{\Omega}^*) - \frac{Np}{2}\log(2\pi) - \frac{1}{2}\left(\tilde{\mathbf{Y}}_i^*\right)'(\boldsymbol{\Omega}^*)^{-1}\left(\tilde{\mathbf{Y}}_i^*\right) \tag{16}$$

where $\boldsymbol{\Omega}^*$ is a tridiagonal matrix with the diagonal elements being $\tilde{\sigma}_i^2\Delta + 2\tilde{a}_i^2$ and the tridiagonal elements being $-\tilde{a}_i^2$, $\tilde{\mathbf{Y}}_i^* = \left(\tilde{Y}_{i,t_1^*} - \tilde{Y}_{i,t_0^*}, ..., \tilde{Y}_{i,t_N^*} - \tilde{Y}_{i,t_{N-1}^*}\right)'$. The QML estimator of $\left(\int_{T-h}^T \tilde{\sigma}_{it}^2 dt, (\mathbf{u}_i^*)'\mathbf{A}_0\mathbf{u}_i^*\right)$ is the value of $(\tilde{\sigma}_i^2, \tilde{a}_i^2)$ which maximizes $l(\tilde{\sigma}_i^2, \tilde{a}_i^2)$. We denote the estimator of $v_i^* = \int_{T-h}^T \tilde{\sigma}_{it}^2 dt$ by $\hat{v}_i^*$, which is positive. Xiu (2010) proved that $\hat{v}_i^*$ is consistent and asymptotically efficient for $\int_{T-h}^T \tilde{\sigma}_{it}^2 dt$.

Remark 3.5. *As discussed in Xiu (2010), if $(t_j^* - t_{j-1}^*)$s for $j = 1, ..., N$ are random and IID, we can add another misspecified assumption that they are equal. We then apply the above approach to get $\hat{v}_i^*$ which is also a consistent estimator of $(\mathbf{u}_i^*)'\boldsymbol{\Sigma}_{T-h,T}\mathbf{u}_i^*$. Since the tick-by-tick data over the time period $[T-h, T]$ is typically non-synchronous, we propose to first synchronize data by* *the refresh time scheme of Barndorff-Nielsen et al. (2011) and then apply the QML procedure to obtain $\hat{v}_i^*$ $(i = 1, \cdots, p)$. The first refresh time t_0^* during a trading day is the first time when all assets have been traded at least once since $T - h$. The second refresh time t_1^* is the first time when all assets have been traded at least once since the first refresh point in time t_0^*. Repeating this sequence yields in total $N+1$ refresh times, $t_0^*, t_1^*, ..., t_N^*$, and corresponding $N+1$ sets of synchronized refresh prices $\mathbf{Y}_{t_0^*}, \mathbf{Y}_{t_1^*}, ..., \mathbf{Y}_{t_N^*}$ with each Y_{i,t_j^*} $(i = 1, ..., p; j = 0, 1, ..., N)$ being the log-price of the ith asset nearest to and previous to t_j^*. Barndorff-Nielsen et al. (2011) showed that if the trading time of p assets arrive as independent standard Poisson processes with common intensity λ such that the mean of trading frequency of each asset over $[T-h, T]$ is λh, then the synchronized data obtained by the refresh time scheme is $\lambda h/\log p$. Based on this observation, if each of 100 (or1,000) assets have around 20,000 observations within a trading day, then the number of synchronized observations is around 4,342 (or 2,895). While this sampling strategy loses around 78.3% or 85.5% of observations, it keeps much more data than the sparsely sampling technique at every 15 minutes, where the size is only 26 within a trading day.*

Therefore, our shrinkage QML estimators for $\mathbf{\Sigma}_{T-h,T}$ and $\mathbf{\Sigma}_{T-h,T}^{-1}$ are, respectively,

$$\widehat{\mathbf{\Sigma}}_{T-h,T} = \mathbf{U}^* \ \mathrm{diag}(\hat{v}_1^*, ..., \hat{v}_p^*) \, (\mathbf{U}^*)', \quad \widehat{\mathbf{\Sigma}_{T-h,T}^{-1}} = \mathbf{U}^* \mathrm{diag} \left\{ (\hat{v}_1^*)^{-1}, ..., (\hat{v}_p^*)^{-1} \right\} (\mathbf{U}^*)', \tag{17}$$

and our estimated optimal weight $\hat{\mathbf{w}}_T$ is obtained by replacing $\mathbf{\Sigma}_{T-h,T}^{-1}$ in (5) with $\widehat{\mathbf{\Sigma}_{T-h,T}^{-1}}$,

$$\hat{\mathbf{w}}_T = \frac{\widehat{\mathbf{\Sigma}_{T-h,T}^{-1}} \mathbf{1}}{\mathbf{1}' \widehat{\mathbf{\Sigma}_{T-h,T}^{-1}} \mathbf{1}}. \tag{18}$$

Notice that like $\mathbf{U}$, $\mathbf{U}^*$ cannot be obtained directly from observations. We therefore approximate $\mathbf{U}^*$ by the eigenvectors of

$$\widetilde{\mathbf{S}}_{0,T-h}^{\mathrm{TVA}} = \frac{\mathrm{tr} \left\{ \sum_{r=1}^m \Delta \mathbf{Y}_r^* \, (\Delta \mathbf{Y}_r^*)' \right\}}{m} \sum_{r=1}^m \frac{\Delta \mathbf{Y}_r^* \, (\Delta \mathbf{Y}_r^*)'}{|\Delta \mathbf{Y}_r^*|^2},$$

where $\Delta \mathbf{Y}_r^* = \mathbf{Y}_{\tau_r^*} - \mathbf{Y}_{\tau_{r-1}^*}$ $(r = 1, ..., m)$, and $\mathbf{Y}_{\tau_r^*}$'s are the log-prices obtained by synchronizing all the trading prices of p assets during $[0, T-h)$ via the previous tick method.

4 Empirical Studies

In this section, we demonstrate the performance of our proposed method using real data. Three portfolio sizes are considered ($p = 30, 40$ and 50) based on stocks traded in the U.S. markets. These portfolios are 30 Dow Jones Industrial Average (30 DJIA) constituent stocks, 30 DJIA stocks and 10 stocks with the largest market caps (ranked on March 30, 2012) from S&P 500 other than 30 DJIA stocks, 30 DJIA stocks and 20 stocks with the largest market caps from S&P 500 other than 30 DJIA stocks. We download daily data starting from March 19, 2012 and ending on December 31, 2013 (450 trading days) from the Center for Research in Security

Prices (CRSP) and 200 days intra-day data staring on March 19, 2013 and ending on December 31, 2013 from the TAQ database. The daily data are used to implement some existing methods in the literature for the purpose of comparison. For the high frequency data, the same data cleaning procedure as in Barndorff-Nielsen et al. (2011) is applied to pre-process the data by 1) deleting entries that have 0 or negative prices, 2) deleting entries with negative values in the column of "Correlation Indicator", 3) deleting entries with a letter code in the column of "COND", except for "E" or "F", 4) deleting entries outside the period 9:30 a.m. to 4 p.m., and 5) using the median price if there are multiple entries at the same time.

4.1 Summary of the proposed method

Given that, in the empirical applications, the basic unit is daily, we can summarize the proposed method as follows. Suppose we want to construct a portfolio strategy at the end of the Jth day (which is denoted T in previous sections) based on a pool of p assets with a holding period of $\breve{J}$ days. We use the ICV in the most recent $J - J_1$ days (which is denoted $[T-h, T]$ in previous sections) multiplied by $\frac{\breve{J}}{J-J_1}$ to approximate the expected ICV during the holding period.

Step 1: Split data of J days into two parts. The first part contains data of first J_1 days, recorded as the 1st, ..., J_1th days. The rest of data of $J - J_1$ days belong to the second part.

Step 2: Synchronize data in the lth day for each $l \in \{1, ..., J_1\}$ using the previous tick method at the 15-minute interval. Denote the log-price at the 15-minute frequency by $\mathbf{Y}_0, \mathbf{Y}_1, ..., \mathbf{Y}_m$.

Step 3: Synchronize the data in lth day for each $l \in \{J_1 + 1, ..., J\}$ using the refresh time scheme to obtain synchronous data and denote the log-price by $\mathbf{Y}^{l*}_{\cdot 0}, \mathbf{Y}^{l*}_{\cdot 1}, ..., \mathbf{Y}^{l*}_{\cdot n_l}$ for each $l \in \{J_1 + 1, ..., J\}$.

Step 4: Obtain the eigenvectors of $\dfrac{\operatorname{tr}\left(\sum_{k=1}^m \Delta\mathbf{Y}_k \Delta\mathbf{Y}_k'\right)}{m} \sum_{k=1}^m \dfrac{\Delta\mathbf{Y}_k \Delta\mathbf{Y}_k'}{|\Delta\mathbf{Y}_k|^2}$ (the corresponding eigenvalues are sorted in the non-increasing order), and put them together as a $p \times p$ matrix which is denoted by $\mathbf{U}^*$. Here $\Delta\mathbf{Y}_k = \mathbf{Y}_k - \mathbf{Y}_{k-1}$.

Step 5: Obtain $\widetilde{\mathbf{Y}}^{l*}_{\cdot j} = (\mathbf{U}^*)' \mathbf{Y}^{l*}_{\cdot j}$ for $l = J_1 + 1, ..., J,\ j = 1, ..., n_l$. Estimate the integrated volatility of the ith element of $(\mathbf{U}^*)' \mathbf{X}_t$ during the lth day by QML that maximizes (16) with $\widetilde{\mathbf{Y}}^*_i$ being replaced by $\widetilde{\mathbf{Y}}^{l*}_{i\cdot} = \left(\tilde{Y}^{l*}_{i1}, ..., \tilde{Y}^{l*}_{i,n_l}\right)$ and with $\tilde{Y}^{l*}_{ij}$ being the ith element of $\widetilde{\mathbf{Y}}^{l*}_{\cdot j}$. Denote the estimator by $\hat{v}^{l*}_i$.

Step 6: The SQML estimator of the ICV in the lth day is defined as $\mathbf{U}^* \operatorname{diag}(\hat{v}^{l*}_1, ..., \hat{v}^{l*}_p) (\mathbf{U}^*)'$. We then use $\frac{\breve{J}}{J-J_1} \sum_{l=J_1+1}^{J} \mathbf{U}^* \operatorname{diag}(\hat{v}^{l*}_1, ..., \hat{v}^{l*}_p) (\mathbf{U}^*)'$ to approximate the expected ICV during the holding period, and its inverse to approximate $\widehat{\mathbf{\Sigma}^{-1}_{T-h,T}}$ in (18) to get the estimated optimal weight.

For the purpose of comparison, we consider two different $\mathbf{U}^*$s. We denote the two different SQML estimators by SQrM if $\mathbf{U}^*$ in Step 4 is obtained from 15-minute intra-day data and SQrD if $\mathbf{Y}_0, ..., \mathbf{Y}_m$ are the daily closing log-prices.

4.2 The GMV portfolio

We first consider the GMV portfolio problem (2) whose theoretical optimal weight is chosen by (3). Following the choice of many practitioners, we apply the plug-in method to estimate the optimal weight and replace $\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}^{-1}$ by its approximation, $\frac{h}{\tau}\widehat{\boldsymbol{\Sigma}_{T-h,T}^{-1}}$ with different hs. We refer to Brandt (2010) for a review of the impacts of a plug-in method in portfolio choice.

We compare the out-of-sample performance of our proposed method with some other methods in the literature, including the equal weight (denoted by EW), the weight estimated by plugging in the optimal linear shrinkage of the sample covariance matrix (denoted by LS), the weight derived by the procedure suggested in Fan, Li and Yu (2012) (denoted by TS). After the weights are determined, the portfolios are constructed accordingly.

LS is obtained by replacing $\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}^{-1}$ in (3) with the inverse of the linear shrinkage estimator

$$\widehat{\mathbf{S}}_{LS} = (1-\kappa)\mathbf{S} + \kappa\bar{\lambda}\mathbf{I}_p,$$

where $\mathbf{S} = J_{LS}^{-1}\sum_{i=1}^{J_{LS}}(\mathbf{Y}_i - \mathbf{Y}_{i-1})(\mathbf{Y}_i - \mathbf{Y}_{i-1})' = \mathbf{Q}\ \mathrm{diag}(\lambda_1, ..., \lambda_p)\mathbf{Q}'$ is the sample covariance matrix of previous J_{LS} daily log-returns, $\lambda_1, ..., \lambda_p$ are the eigenvalues of $\mathbf{S}$, $\mathbf{Q}$ contains corresponding eigenvectors, $\bar{\lambda} = \sum_{i=1}^{p}\lambda_i/p$, and κ is determined by the asymptotic optimization results derived in Ledoit and Wolf (2004).

Fan, Li and Yu (2012) considered the following risk optimization problem under gross-exposure constraints

$$\min \mathbf{w}'\boldsymbol{\Sigma}_{T-h,T}\mathbf{w} \quad s.t. \quad \|\mathbf{w}\|_1 \le c \quad \text{and} \quad \mathbf{w}'\mathbf{1} = 1, \tag{19}$$

where $\boldsymbol{\Sigma}_{T-h,T}$ was also used to approximate $\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}$. The pair-wise two scales covariance (TSCV) estimator of $\boldsymbol{\Sigma}_{T-h,T}$ was constructed based on the high frequency data synchronized by the pair-wise refresh time scheme over previous J_{TS} trading days. Since this pair-wise estimator may not be positive semi-definite, they projected the estimator (denoted by $\mathbf{M}$ here) by

$$\mathbf{M}_1 = (\mathbf{M} + \lambda_{\min}^{-}\mathbf{I}_p)/(1 + \lambda_{\min}^{-}), \tag{20}$$

where $\lambda_{\min}^{-}$ is the negative part of the minimum eigenvalue of the estimator $\mathbf{M}$. They then minimize $\mathbf{w}'\mathbf{M}_1\mathbf{w}$ to obtain the optimal weight $\widehat{\mathbf{w}}$ for a given c. In this paper, following the simulation and the empirical studies in Fan, Li and Yu (2012), we set $c = 1.2$.

In practice, one choice that we have to make is the number of days over which we do the estimation. For our new developed approach, we let $J_1 = 50, 60, ..., 250$ when we use daily log-returns, and let $J_1 = 5$ (one week), 6, ..., 21 (one month) days when we use 15-minute intra-day log-returns in Step 4. Moreover, we choose $J - J_1 = 1, 2, ..., 5$. The optimal result among all possible combinations is reported. Similarly, we report the optimal results for LS when $J_{LS} \in \{50, 60, ..., 250\}$ and TS when $J_{TS} \in \{1, 2, ..., 10\}$, and denote them by TSo, LSo respectively.

The following three measures are calculated to compare the out-of-sample performance of all the methods during 174 investment days (we have 200 days intra-day data in total and we use

26 days intra-day data to get SQrM), from April 25, 2013 to December 31, 2013: (1) the average of log-returns of the portfolio multiplied by 252 (denoted by AV); (2) the standard deviation of log-returns of the portfolio multiplied by $\sqrt{252}$ (denoted by SD); (3) information ratio calculated by AV/SD (denoted by IR).

In general, a high AV and a high IR with a low SD are expected for a good portfolio. Since the GMV portfolio is designed to minimize the variance of a portfolio, the most important performance measure for GMV is SD. Therefore, we first compare the standard deviations of different methods and then compare the information ratios and the average returns.

Reported in Table 1 are the AV, SD and IR for all the methods. The number in the bold face represents the lowest SD. Several conclusions can be made from Table 1. First and foremost, SQrM outperforms all the other strategies in terms of SD. SQrM also achieves the highest information ratio when $p = 50$. Second, as expected, the standard deviation of the GMV portfolio decreases, as p increases from 30 to 50, for most methods. The only exception is the EW. Third, SQrM performs better than SQrD, indicating that high frequency data are useful in portfolio choice.

4.3 Markowitz portfolio with momentum signals (MwM)

We now consider a 'full' Markowitz portfolio without any short-sale constraint. The Markowitz portfolio minimizes the variance of a portfolio under two conditions:

$$\min \mathbf{w}'\widetilde{\boldsymbol{\Sigma}}_{T,T+\tau}\mathbf{w} \quad \text{subject to} \quad \mathbf{w}'\mathbf{1} = 1 \quad \text{and} \quad \mathbf{w}'\mathbf{e} = b,$$

where b is a target expected return chosen by an investor and $\mathbf{e}$ is a signal to denote the vector of expected returns of p assets. The above problem has the following analytical solution

$$\mathbf{w} = c_1\widetilde{\boldsymbol{\Sigma}}^{-1}_{T,T+\tau}\mathbf{1} + c_2\widetilde{\boldsymbol{\Sigma}}^{-1}_{T,T+\tau}\mathbf{e}, \tag{21}$$

where

$$c_1 = \frac{C - bB}{AC - B^2}, \quad c_2 = \frac{bA - B}{AC - B^2}, \quad A = \mathbf{1}'\widetilde{\boldsymbol{\Sigma}}^{-1}_{T,T+\tau}\mathbf{1}, \quad B = \mathbf{1}'\widetilde{\boldsymbol{\Sigma}}^{-1}_{T,T+\tau}\boldsymbol{e}, \quad C = \boldsymbol{e}'\widetilde{\boldsymbol{\Sigma}}^{-1}_{T,T+\tau}\boldsymbol{e}.$$

To choose $\mathbf{e}$ and b, we follow Ledoit and Wolf (2014). In particular, the ith element of $\mathbf{e}$ is the momentum factor which is chosen as the arithmetic average of the previous 250 days returns on the ith stock. b is the arithmetic average of the momentums of the top-quintile stocks according to $\mathbf{e}$. In Table 2, we report the annualized AV, SD, and IR of the daily log-returns for all methods, namely, SQrM, SQrD, LS, the equal weight constructed on top-quintile stocks according to their momentums (denoted by EW-TQ), and the method with $\widetilde{\boldsymbol{\Sigma}}^{-1}_{T,T+\tau}$ in (21) being replaced by the inverse of the sample covariance matrix of previous J_{SP} days daily log-returns (denoted by SP when $J_{SP} = 250$). Similar to the GMV portfolio, we choose optimal J and J_1 for SQrM and SQrD. For the Markowitz portfolio, a more relevant criterion for the comparison is IR. In this paper, we first compare the IRs and then the SDs.

In Table 2, the number in bold face represents the highest IR while the number with a '*' represents the lowest SD. It can be seen that SQrM and SQrD perform better than other methods in terms of IR except the EW when $p = 30, 40$. However, the SDs of SQrM and SQrD are much lower than that of EW and also lower than that of the other methods.

4.4 Robustness of sample period

To check the robustness of our strategy, we split the entire 174 investment days into two subperiods, one from April 25, 2013 to August 27, 2013 and the other from August 28, 2013 to December 31, 2013. The results of the GMV portfolio are reported in Tables 3 and 4. It can be seen that SQrM and SQrD continue to outperform other methods in terms of SD in all cases. Empirical results of the Markowitz portfolio with the momentum signal are reported in Tables 5 and 6. Again SQrM and SQrD continue to outperform other methods in almost all cases in terms of IR and SD.

We also perform a moving-window analysis to check the robustness of our empirical results. Staring from April 25, 2013, we calculate the standard deviation of daily log-returns of each method over 42 trading days and repeat this exercise by moving one trading day at each pass. To compare our method with other methods, we use figures to show the results of TSo, LSo, SQrD and SQrM for the GMV portfolio and SPo, LSo, SQrD and SQrM for the MwM portfolio, where the optimal numbers of days chosen for each method is to minimize or maximize the mean of SDs or IRs of 133 different investment periods (each investment period is 42 days) for the GMV and MwM portfolio, respectively. Figures 1, 2, 3 plot the results when $p = 30$, 40 and 50.

We find that SQrM performs better and better as the portfolio size increases. In general, it has the lowest SD and the highest IR for both the GMV portfolio and the MwM portfolios. This result indicates that high frequency data are useful in portfolio choice, especially for controlling the risk.

4.5 Robustness of time span

From a statistical perspective, a longer span of historical data contains more information about the dynamic of an asset price so that it may be reasonable to believe that methods based on a longer span of data should perform better than those based on less data. However, the model specification is more likely to be wrong over a longer span. Hence there is a trade off between the estimation error and the specification error. In this subsection we examine this trade off empirically in the context of the LS portfolio and the SQrD portfolio. In particular, the LS portfolio and the SQrD portfolio are constructed based on different historical data sets for the GMV portfolio. For LS, we set $J_{LS} = 50, 60, ..., 250$. For SQrD, we fix $J - J_1 = 1$ and set $J_1 = 50, 60, ..., 250$.

Figure 4 plots the risk of daily log-returns for the two GMV portfolios as a function of J_{LS} or J_1 when $p = 30$ and $p = 50$. Some interesting findings emerge. First, the risk of log-returns of a portfolio does not necessarily decrease when a longer span of historical data is used. Second, SQrD performs better than LS in almost all cases and is more stable across different time spans. This is especially true when $p = 50$. Once again, there is an advantage for using our estimator for portfolio selection. Third, when $p = 30$, the risk of the SQrD portfolio decreases when J_1 increases initially. This is because more data are used in estimation, reducing the estimation error. However, the risk increases when $J_1 > 110$. This is because the construction of SQrD relies on the assumption that $\mathbf{X}_t \in \mathcal{C}$. As J_1 increases, the time span becomes longer, and hence the assumption that $\mathbf{X}_t \in \mathcal{C}$ is more likely to be invalid. This can also explain why SQrM performs better than SQrD.

5 Conclusions

This paper has developed a new estimator for the ICV and its inverse from high frequency data when the portfolio size p and the sample size of data n satisfies $p/n \to y > 0$ as n goes to ∞. The use of high frequency data drastically increases the sample size and hence reduces the estimation error. To further prevent the estimation error from accumulating with p, a new regularization method is applied to the eigenvalues of an initial estimator of the ICV. Our proposed estimator of the ICV is always positive definite and its inverse is the estimator of the inverse of the ICV. It minimizes the limit of the out-of-sample variance of portfolio returns within the class of rotation-equivalent estimators. It works when the number of underlying assets is larger than the number of time series observations in each asset and when the asset price follows a general stochastic process.

The asymptotic optimality for our proposed method is justified under the assumption that $p/n \to y > 0$ as n goes to ∞. The usefulness of our estimator is examined in real data. The method is used to construct the optimal weight in the global minimum variance and the Markowitz portfolio with momentum signal based on the DJIA 30 and another 20 stocks chosen from S&P500. The performance of our proposed method is compared with that of some existing methods in the literature. The empirical results show that our method performs favorably out-of-sample.

6 Appendix

In the appendix we first prove Theorem 3.2 as the proof of Theorem 3.1 relies on Theorem 3.2.

Proof of Theorem 3.2. By assumption (A.i), we can write

$$\Delta \mathbf{X}_k = \int_{\tau_{k-1}}^{\tau_k} \gamma_t \mathbf{\Lambda} d\mathbf{W}_t \overset{d}{=} \left(\int_{\tau_{k-1}}^{\tau_k} \gamma_t^2 dt \right)^{1/2} \breve{\mathbf{\Sigma}}^{1/2} \mathbf{z}_k,$$

where '$\overset{d}{=}$' stands for 'equal in distribution', $\breve{\mathbf{\Sigma}} = \mathbf{\Lambda}\mathbf{\Lambda}'$ and $\mathbf{z}_k = (Z_{1k}, \cdots, Z_{pk})'$ consists of independent standard normals. Then

$$\mathbf{S}_{T-h,T}^{\mathrm{TVA}} = \frac{\mathrm{tr}\left(\widehat{\mathbf{\Sigma}}_{T-h,T}^{\mathrm{RCV}}\right)}{p} \cdot \frac{p}{n} \sum_{k=1}^{n} \breve{\mathbf{\Sigma}}^{1/2} \left(\frac{\mathbf{z}_k \mathbf{z}_k'}{\mathbf{z}_k' \breve{\mathbf{\Sigma}} \mathbf{z}_k} \right) \breve{\mathbf{\Sigma}}^{1/2},$$

where $\widehat{\boldsymbol{\Sigma}}^{\mathrm{RCV}}_{T-h,T} = \sum_{k=1}^{n} \Delta\mathbf{X}_k \Delta\mathbf{X}_k'$, $\Delta\mathbf{X}_k = \mathbf{X}_{\tau_k} - \mathbf{X}_{\tau_{k-1}}$. Denote

$$\mathbf{S}^{IID}_{T-h,T} := \sum_{k=1}^{n} \frac{1}{n} \boldsymbol{\Sigma}^{1/2}_{T-h,T} \mathbf{z}_k \mathbf{z}_k' \boldsymbol{\Sigma}^{1/2}_{T-h,T} = \int_{T-h}^{T} \gamma_t^2 dt \cdot \left(\frac{1}{n} \sum_{k=1}^{n} \breve{\boldsymbol{\Sigma}}^{1/2} \mathbf{z}_k \mathbf{z}_k' \breve{\boldsymbol{\Sigma}}^{1/2} \right).$$

From Theorem 2 of Ledoit and Pèchè (2011), we know that $p^{-1}\mathrm{tr}\left\{ \left(\mathbf{S}^{IID}_{T-h,T} - z\mathbf{I}\right)^{-1} \boldsymbol{\Sigma}_{T-h,T} \right\}$ converges to

$$s_\Psi(z) = \int \frac{r}{r\{1 - y - yz \times s_F(z)\} - z} \, dH(r),$$

almost surely, where H is the LSD of matrices $\boldsymbol{\Sigma}_{T-h,T}$ and F is the LSD of matrices $\mathbf{S}^{IID}_{T-h,T}$ or $\mathbf{S}^{\mathrm{TVA}}_{T-h,T}$, since they share the same LSD by Theorem 2 of Zheng and Li (2011).

On the other hand, we have that $s_\Psi(z)$ is the Stieltjes transform of the bounded function $\Psi(x)$ defined in (13) by Theorem 4 of Ledoit and Pèchè (2011) and the Stieltjes transform of function $\Psi_p(x)$ is

$$s_{\Psi_p}(z) = \frac{1}{p} \mathrm{tr}\left\{ \left(\mathbf{S}^{\mathrm{TVA}}_{T-h,T} - z\mathbf{I}\right)^{-1} \boldsymbol{\Sigma}_{T-h,T} \right\}.$$

Therefore we only need to show that

$$\frac{1}{p} \mathrm{tr}\left\{ \left(\mathbf{S}^{IID}_{T-h,T} - z\mathbf{I}\right)^{-1} \boldsymbol{\Sigma}_{T-h,T} \right\} - \frac{1}{p} \mathrm{tr}\left\{ \left(\mathbf{S}^{\mathrm{TVA}}_{T-h,T} - z\mathbf{I}\right)^{-1} \boldsymbol{\Sigma}_{T-h,T} \right\} \stackrel{a.s.}{\to} 0.$$

To prove this, it suffices to show the following two facts:

$$\max_{1\le k\le n} \left| \frac{1}{p} \mathbf{z}_k' \breve{\boldsymbol{\Sigma}} \mathbf{z}_k - 1 \right| \stackrel{a.s.}{\to} 0, \tag{22}$$

and

$$\frac{1}{p} \mathrm{tr}\left(\widehat{\boldsymbol{\Sigma}}^{\mathrm{RCV}}_{T-h,T} \right) - \int_{T-h}^{T} \gamma_t^2 dt \stackrel{a.s.}{\to} 0. \tag{23}$$

To prove (22), by assumption (A.iii), all the eigenvalues of $\breve{\boldsymbol{\Sigma}}$ are bounded, so that $\mathrm{tr}(\breve{\boldsymbol{\Sigma}}^r) = O(p)$ for all $1 \le r < \infty$. From Lemma 2.7 of Bai and Silverstein (1998), we have

$$\begin{aligned}
\mathrm{E}\left(\max_{1\le k\le n} \left| p^{-1} \mathbf{z}_k' \breve{\boldsymbol{\Sigma}} \mathbf{z}_k - 1 \right|^6 \right) &\le \frac{Cn}{p^6} \left(\left\{ \mathrm{E}|Z_{jk}|^4 \mathrm{tr}\left(\breve{\boldsymbol{\Sigma}}^2\right) \right\}^3 + \mathrm{E}|Z_{jk}|^{12} \mathrm{tr}\left(\breve{\boldsymbol{\Sigma}}^6\right) \right) \\
&= O(n^{-2}),
\end{aligned}$$

where the last step comes from the fact that the higher order moments of Z_{jk}'s are finite since they are normally distributed. Thus, (22) follows by the Borel-Cantelli lemma.

We now prove (23).

$$\begin{aligned}
\left| p^{-1} \mathrm{tr}\left(\widehat{\boldsymbol{\Sigma}}^{\mathrm{RCV}}_{T-h,T} \right) - \int_{T-h}^{T} \gamma_t^2 dt \right| &= \left| p^{-1} \sum_{k=1}^{n} \int_{\tau_{k-1}}^{\tau_k} \gamma_t^2 dt \cdot \mathbf{z}_k' \breve{\boldsymbol{\Sigma}} \mathbf{z}_k - \int_{T-h}^{T} \gamma_t^2 dt \right| \\
&= \left| \sum_{k=1}^{n} \int_{\tau_{k-1}}^{\tau_k} \gamma_t^2 dt \cdot \left(p^{-1} \mathbf{z}_k' \breve{\boldsymbol{\Sigma}} \mathbf{z}_k - 1 \right) \right| \\
&\le \max_{1\le k\le n} \left| p^{-1} \mathbf{z}_k' \breve{\boldsymbol{\Sigma}} \mathbf{z}_k - 1 \right| \cdot \int_{T-h}^{T} \gamma_t^2 dt \\
&\stackrel{a.s.}{\to} 0.
\end{aligned}$$

by assumption (A.iv) and the result in Equation (22).

Then,

$$\begin{aligned}
& \frac{1}{p}\mathrm{tr}\left\{\left(\mathbf{S}^{IID}_{T-h,T}-z\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\right\}-\frac{1}{p}\mathrm{tr}\left\{\left(\mathbf{S}^{\mathrm{TVA}}_{T-h,T}-z\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\right\} \\
= \; & \frac{1}{p}\mathrm{tr}\left\{\left(\mathbf{S}^{IID}_{T-h,T}-z\mathbf{I}\right)^{-1}\left(\mathbf{S}^{\mathrm{TVA}}_{T-h,T}-\mathbf{S}^{IID}_{T-h,T}\right)\left(\mathbf{S}^{\mathrm{TVA}}_{T-h,T}-z\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\right\} \\
= \; & \frac{1}{p}\mathrm{tr}\left\{\left(\mathbf{S}^{IID}_{T-h,T}-z\mathbf{I}\right)^{-1}\frac{\mathrm{tr}\left(\widehat{\boldsymbol{\Sigma}}^{\mathrm{RCV}}_{T-h,T}\right)}{p(n)}\cdot\sum_{k=1}^{n}\left(\frac{1}{p^{-1}\mathbf{z}_k'\breve{\boldsymbol{\Sigma}}\mathbf{z}_k}-1\right)\breve{\boldsymbol{\Sigma}}^{1/2}\mathbf{z}_k\mathbf{z}_\ell'\breve{\boldsymbol{\Sigma}}^{1/2}\left(\mathbf{S}^{\mathrm{TVA}}_{T-h,T}-z\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\right\} \\
& +\frac{1}{p(n)}\mathrm{tr}\left(\left(\mathbf{S}^{IID}_{T-h,T}-z\mathbf{I}\right)^{-1}\left\{p^{-1}\mathrm{tr}\left(\widehat{\boldsymbol{\Sigma}}^{\mathrm{RCV}}_{T-h,T}\right)-\int_{T-h}^{T}\gamma_t^2dt\right\}\right. \\
& \left.\times\sum_{k=1}^{n}\breve{\boldsymbol{\Sigma}}^{1/2}\mathbf{z}_k\mathbf{z}_k'\breve{\boldsymbol{\Sigma}}^{1/2}\left(\mathbf{S}^{\mathrm{TVA}}_{T-h,T}-z\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\right) \\
:= \; & I_1+I_2.
\end{aligned}$$

From assumptions (A.ii)-(A.v), and the facts that $\|(\mathbf{S}^{IID}_{T-h,T}-z\mathbf{I})^{-1}\|\le 1/\Im(z)$, $\|(\mathbf{S}^{\mathrm{TVA}}_{T-h,T}-z\mathbf{I})^{-1}\|\le 1/\Im(z)$ with $\|\cdot\|$ denoting the L_2 norm of a matrix, (22) and (23), we have that both $|I_1|$ and $|I_2|$ converge to 0, almost surely. Therefore, the proof of Theorem 3.2 is completed. □

Proof of Theorem 3.1. The convergence of ESD of $\mathbf{S}^{\mathrm{TVA}}_{T-h,T}$ is shown in Theorem 2 of Zheng and Li (2011). Note that $\left\|\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\right\|\le C$ for some fixed number C when p large enough by assumption (A.v) and the fact that $\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}$ belongs to class $\mathcal{S}$. Thus, from Lemma 2.7 of Bai and Silverstein (1998) and Borel-Cantelli lemma, we have

$$\frac{1}{p}\mathbf{1}'\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\mathbf{1}-\frac{1}{p}\mathrm{tr}\left\{\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\right\}\xrightarrow{a.s.}0.$$

Moreover, we have

$$\begin{aligned}
\frac{1}{p}\mathrm{tr}\left\{\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\right\} &= \frac{1}{p}\sum_{i=1}^{p}\frac{1}{g_n(v_i)}=\int\frac{1}{g_n(x)}dF^{\mathbf{S}^{\mathrm{TVA}}_{T-h,T}}(x) \\
&\xrightarrow{a.s.}\int\frac{1}{g(x)}dF(x).
\end{aligned}$$

Therefore,

$$\frac{1}{p}\mathbf{1}'\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\mathbf{1}\xrightarrow{a.s.}\int\frac{1}{g(x)}dF(x). \tag{24}$$

Similarly, we can show that

$$\frac{1}{p}\mathbf{1}'\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\mathbf{1}-\frac{1}{p}\mathrm{tr}\left\{\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\right\}\xrightarrow{a.s.}0.$$

Using Theorem 3.2, we have

$$
\begin{aligned}
\frac{1}{p}\mathrm{tr}\left\{\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\right\} &= \frac{1}{p}\mathrm{tr}(\mathbf{U}'\boldsymbol{\Sigma}_{T-h,T}\mathbf{U}\mathbf{V}^{-2}) = \frac{1}{p}\sum_{i=1}^{p}\frac{\mathbf{u}_i'\boldsymbol{\Sigma}_{T-h,T}\mathbf{u}_i}{g_n(v_i)^2} \\
&\xrightarrow{a.s.} \int\frac{x}{|1-y-yx\times\breve{m}_F(x)|^2 g(x)^2}dF(x).
\end{aligned}
$$

Thus,

$$
\frac{1}{p}\mathbf{1}'\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\boldsymbol{\Sigma}_{T-h,T}\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\mathbf{1}\xrightarrow{a.s.}\int\frac{x}{|1-y-yx\times\breve{m}_F(x)|^2 g(x)^2}dF(x). \tag{25}
$$

Combining (24) and (25), we obtain that

$$
p\cdot\frac{\mathbf{1}'\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\boldsymbol{\Sigma}_{0,T-h}\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\mathbf{1}}{\left(\mathbf{1}'\left(\widehat{\boldsymbol{\Sigma}}^{*}_{T-h,T}\right)^{-1}\mathbf{1}\right)^2}\xrightarrow{a.s.}\frac{\int\frac{x}{|1-y-yx\times\breve{m}_F(x)|^2 g(x)^2}dF(x)}{\left(\int\frac{dF(x)}{g(x)}\right)^2}.
$$

□

References

[1] Abadir, K. M., Distaso, W., Žikěs, F., 2014. Design-free estimation of variance matrices. Journal of Econometrics 181(2), 165-180.

[2] Aït-Sahalia, Y., Fan, J., Xiu, D., 2010. High frequency covariance estimates with noisy and asynchronous financial data. Journal of the American Statistical Association 105, 1504-1517.

[3] Aït-Sahalia, Y., Xiu, D., 2016. A Hausman test for the presence of market microstructure noise in high frequency data. Working Paper.

[4] Bai, Z., Silverstein, J., 1998. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. The Annals of Probability 26(1), 316-345.

[5] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shepard, 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, 149-169.

[6] Brandt, M. W., 2010. Portfolio choice problems in Y. Aït-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, Volume 1: Tools and Techniques, North Holland, 269-336.

[7] DeMiguel, V., Garlappi, L., Nogales, F. J., Uppal, R., 2009. A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Management Science 55(5), 798-812.

[8] DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? Review of Financial Studies 22, 1915-1953.

[9] Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. Journal of Econometrics 147, 186-197.

[10] Fan, J., Li, Y., Yu, K., 2012. Vast volatility matrix estimation using high-frequency data for portfolio selection. Journal of the American Statistical Association 107, 412-428.

[11] Frahm, G., Memmel, C., 2010. Dominating estimators for minimum-variance portfolios. Journal of Econometrics 159(2), 289-302.

[12] Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. Journal of Finance 58(4), 1651-1684.

[13] Jobson, J., Korkie, B., 1980. Estimation for Markowitz efficient portfolios. Journal of the American Statistical Association 75, 544-554.

[14] Kan, J., Zhou, G., 2007. Optimal portfolio choice with parameter uncertainty. Journal of Financial and Quantitative Analysis 42, 621-656.

[15] Lam, C., 2016. Nonparametric eigenvalue-regularized precision or covariance matrix estimator. The Annals of Statistics 44(3), 928-953.

[16] Ledoit, O., Pèchè, S., 2011. Eigenvectors of some large sample covariance matrix ensembles. Probability Theory and Related Fields 151(1-2), 233-264.

[17] Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10(5), 603-621.

[18] Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88(2), 365-411.

[19] Ledoit, O., Wolf, M., 2014. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. Working Paper.

[20] Liu, C., Tang, C. Y., 2014. A quasi-maximum likelihood approach for integrated covariance matrix estimation with high frequency data. Journal of Econometrics 180, 217-232.

[21] Markowitz, H., 1952. Portfolio selection. Journal of Finance 7, 77-91.

[22] Michaud, R., 1989. The Markowitz optimization enigma: is optimization optimal? Financial Analysts Journal 45(1), 31.

[23] Muirhead, R. J., 1987. Developments in eigenvalue estimation. *In Advances in Multivariate Statistical Analysis* (A. K. Gupta, ed.) 277-288. Reidel, Dordrecht.

[24] Pesaran, M. H., Zaffaroni, P., 2009. Optimality and diversifiability of mean variance and arbitrage pricing portfolios. Working Paper.

[25] Stein, C., 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 197-206. University of California Press.

[26] Stein, C., 1975. Estimation of a covariance matrix. Rietz lecture, 39th Annual Meeting IMS. Atlanta, Georgia.

[27] Tu, J., Zhou, G., 2011. Markowitz meets Talmud: a combination of sophisticated and naive diversification strategies. Journal of Financial Economics 99, 204-215.

[28] Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. Journal of Econometrics 159, 235-250.

[29] Zhang, L., 2011. Estimating covariation: epps effect, microstructure noise. Journal of Econometrics 160, 33-47.

[30] Zheng, X., Li, Y., 2011. On the estimation of integrated covariance matrices of high dimensional diffusion process. The Annals of Statistics 39, 3121-3151.

Table 1: The out-of-sample performance of different daily rebalanced strategies for the GMV portfolio between April 25, 2013 and December 31, 2013.

Period: 04/25/2013—12/31/2013							
$p = 30$	EW	TS	TSo	LS	LSo	SQrD	SQrM
AV	20.13	13.22	13.22	15.31	12.96	10.59	15.62
SD	10.17	9.65	9.65	9.80	9.52	9.34	**9.17**
IR	1.98	1.37	1.37	1.56	1.36	1.80	1.70
$p = 40$	EW	TS	TSo	LS	LSo	SQrD	SQrM
AV	21.00	16.51	17.80	16.00	11.84	19.06	18.09
SD	10.43	9.66	9.62	9.85	9.29	9.29	**9.10**
IR	2.01	1.71	1.85	1.62	1.27	2.05	1.99
$p = 50$	EW	TS	TSo	LS	LSo	SQrD	SQrM
AV	21.00	20.15	20.15	13.28	10.25	17.74	20.52
SD	10.36	9.40	9.40	9.47	9.18	9.26	**8.68**
IR	2.03	2.14	2.14	1.40	1.12	1.91	2.36

Note: AV, SD, IR denote the average, standard deviation, and information ratio of 174 daily log-returns, respectively. AV and SD are annualized and in percent. The smallest number in the row labeled by SD is reported in bold face. TSo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the two-scale covariance matrix obtained based on historical intra-day data (10 days when $p = 30, 50$; 8 days when $p = 40$). LSo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the linear shrinkage of the sample covariance matrix of daily log-returns (110, 90 and 90 days when $p = 30, 40$ and 50 respectively). The optimal number of days is chosen by minimizing SD of 174 log-returns of each portfolio.

Table 2: The out-of-sample performance of different daily rebalanced strategies for Markowitz portfolio with momentum signal between April 25, 2013 and December 31, 2013.

Period: 04/25/2013—12/31/2013							
$p = 30$	EW-TQ	SP	SPo	LS	LSo	SQrD	SQrM
AV	31.74	0.02	4.18	13.02	13.02	20.02	15.91
SD	13.27	12.10	12.06	11.59	11 56	11.37	11.11*
IR	**2.39**	0.00	0.35	1.12	1.12	1.76	1.43
$p = 40$	EW	SP	SPo	LS	LSo	SQrD	SQrM
AV	36.49	6.86	8.94	18.98	18.98	24.67	20.21
SD	13.55	10.90	10.99	11.29	10.66	11.25	10.07*
IR	**2.69**	0.63	0.81	1.68	1.68	2.19	2.01
$p = 50$	EW-TQ	SP	SPo	LS	LSo	SQrD	SQrM
AV	28.64	7.53	11.33	15.81	15.81	22.65	23.35
SD	13.16	10.42	10.80	10.60	10.60	10.51	9.83*
IR	2.18	0.72	1.05	1.49	2.03	2.15	**2.38**

Note: AV, SD, IR denote the average, standard deviation, and information ratio of 174 daily log-returns respectively. AV, SD are annualized and in percent. The smallest number in the row labeled by SD is reported in bold face. SPo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the sample covariance matrix of daily log-returns (190, 230 and 130 days when $p = 30, 40$ and 50, respectively). LSo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the linear shrinkage of the sample covariance matrix of daily log-returns (250 days when $p = 30, 40$ and 50). The optimal number of days is chosen by maximizing IR of 174 log-returns of each portfolio.

Table 3: The out-of-sample performance of different daily rebalanced strategies for the GMV portfolio between April 25, 2013 and August 27, 2013.

Period: 04/25/2013—08/27/2013							
$p = 30$	EW	TS	TSo	LS	LSo	SQrD	SQrM
AV	5.23	2.23	2.23	2.63	9.90	2.93	1.87
SD	10.93	10.05	10.05	10.64	9.85	9.97	**9.81**
IR	0.48	0.22	0.22	0.25	1.00	0.29	0.19
$p = 40$	EW	TS	TSo	LS	LSo	SQrD	SQrM
AV	5.04	3.64	4.35	0.48	4.49	0.43	0.42
SD	11.09	10.26	10.00	10.51	9.45	9.78	**9.73**
IR	0.45	0.36	0.43	0.05	0.48	0.04	0.04
$p = 50$	EW	TS	TSo	LS	LSo	SQrD	SQrM
AV	6.53	9.35	9.35	-1.78	8.29	1.70	5.70
SD	11.06	10.12	10.12	10.18	**9.18**	9.89	9.30
IR	0.59	0.92	0.92	-0.17	0.89	0.17	0.61

Note: AV, SD, IR denote the average, standard deviation, and information ratio of 87 daily log-returns, respectively. AV, SD are annualized and in percent. The smallest number in the row labeled by SD is reported in bold face. TSo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the two-scale covariance matrix obtained based on historical intra-day data (10 days when $p = 30, 50$; 8 days when $p = 40$). LSo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the linear shrinkage of the sample covariance matrix of daily log-returns (110, 90 and 90 days when $p = 30, 40$ and 50 respectively).

Table 4: The out-of-sample performance of different daily rebalanced strategies for the GMV portfolio between August 28, 2013 to December 31, 2013.

Period: 08/28/2013—12/31/2013							
$p = 30$	EW	TS	TSo	LS	LSo	SQrD	SQrM
AV	35.04	24.22	24.22	27.99	16.02	30.70	29.36
SD	9.31	9.24	9.24	8.87	9.17	8.64	**8.45**
IR	3.76	2.62	2.62	3.15	1.75	3.55	3.48
$p = 40$	EW	TS	TSo	LS	LSo	SQrD	SQrM
AV	36.97	29.37	31.25	31.52	19.19	20.15	25.94
SD	9.69	9.01	9.19	9.09	9.16	8.54	**8.32**
IR	3.82	3.26	3.40	3.47	2.09	4.411	4.30
$p = 50$	EW	TS	TSo	LS	LSo	SQrD	SQrM
AV	35.47	30.95	30.95	28.33	12.20	33.78	35.34
SD	9.60	8.64	8.64	8.67	9.05	8.52	**7.95**
IR	3.70	3.58	3.58	3.27	1.35	3.96	4.44

Note: AV, SD, IR denote the average, standard deviation, and information ratio of 87 daily log-returns, respectively. AV, SD are annualized and in percent. The smallest number in the row labeled by SD is reported in bold face. TSo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the two-scale covariance matrix obtained based on historical intra-day data (10 days when $p = 30, 50$ and 8 days when $p = 40$). LSo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the linear shrinkage of the sample covariance matrix of daily log-returns (110, 90 and 90 days when $p = 30, 40$ and 50 respectively).

Table 5: The out-of-sample performance of different daily rebalanced strategies for the Markowitz portfolio with momentum signal between April 25, 2013 to August 27, 2013.

Period: 04/25/2013—08/27/2013							
$p = 30$	EW-TQ	SP	SPo	LS	LSo	SQrD	SQrM
AV	8.70	-10.05	-1.43	-2.99	-2.99	1.73	0.13
SD	14.51	13.02	13.03	12.54	12.54	11.94*	12.06
IR	**0.60**	-0.81	-0.11	-0.24	-0.24	0.01	0.04
$p = 40$	EW-TQ	SP	SPo	LS	LSo	SQrD	SQrM
AV	9.21	-11.39	-9.39	-4.04	-4.04	4.14	-2.69
SD	14.19	10.91	10.94	11.45	10.99	11.49	10.42*
IR	**0.65**	-1.04	-0.86	-0.35	-0.35	0.36	-0.26
$p = 50$	EW-TQ	SP	SPo	LS	LSo	SQrD	SQrM
AV	3.50	-8.15	1.71	-6.25	-6.25	6.98	4.16
SD	14.28	11.01	10.48	10.92	10.92	10.84	10.47*
IR	0.25	-0.74	0.16	-0.57	-0.57	**0.64**	0.40

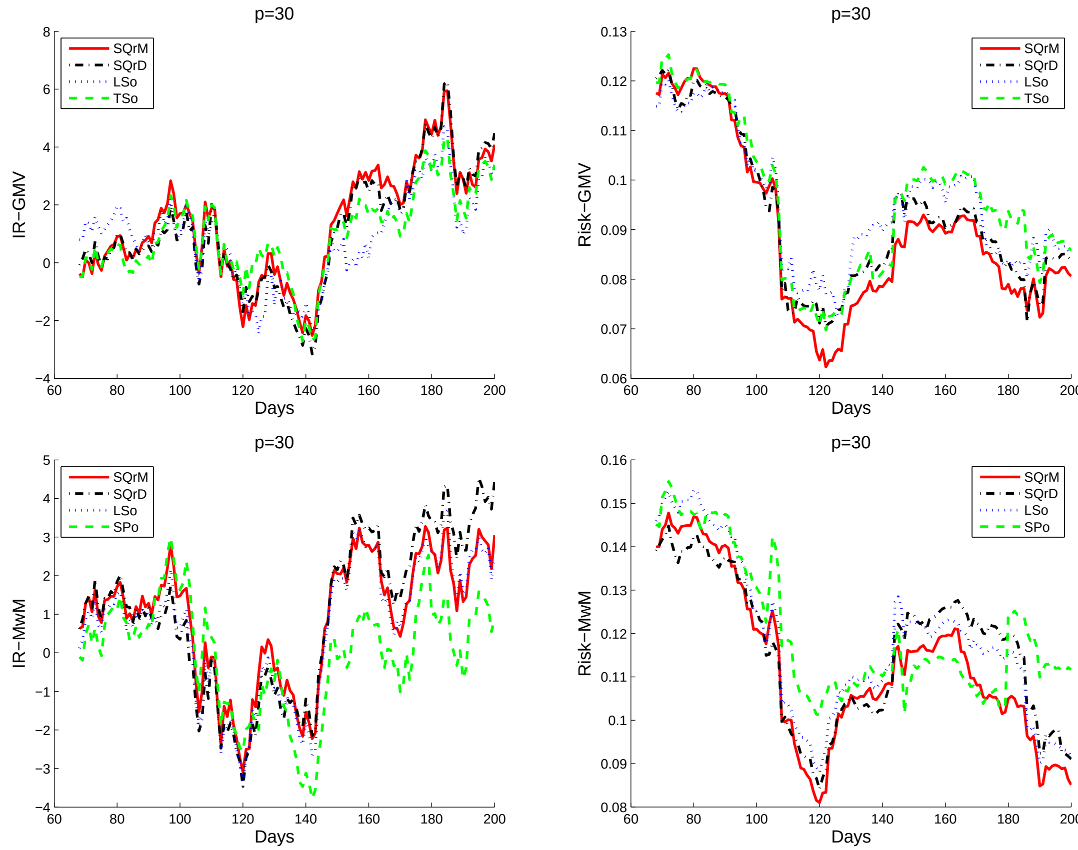
Note: AV, SD, IR denote the average, standard deviation, and information ratio of 87 daily log-returns respectively. AV, SD are annualized and in percent. The smallest number in the row labeled by SD is reported in bold face. SPo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the sample covariance matrix of daily log-returns (190, 230 and 130 days when $p = 30, 40$ and 50 respectively). LSo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the linear shrinkage of the sample covariance matrix of daily log-returns (250 days when $p = 30, 40$ and 50).

Table 6: The out-of-sample performance of different daily rebalanced strategies for the Markowitz portfolio with momentum signal between August 28, 2013 to December 31, 2013.

Period: 08/28/2013—12/31/2013							
$p = 30$	EW-TQ	SP	SPo	LS	LSo	SQrD	SQrM
AV	54.79	10.59	9.79	29.02	29.02	39.95	31.35
SD	11.82	11.14	11.06	10.53	10.53	10.69	10.02*
IR	**4.64**	0.95	0.89	2.76	2.76	3.74	3.11
$p = 40$	EW-TQ	SP	SPo	LS	LSo	SQrD	SQrM
AV	63.78	25.11	27.28	41.99	41.99	45.21	43.11
SD	12.73	10.84	10.97	11.00	11.00	10.92	9.55*
IR	**5.01**	2.32	2.49	3.82	3.82	4.14	4.51
$p = 50$	EW-TQ	SP	SPo	LS	LSo	SQrD	SQrM
AV	53.77	23.20	20.94	37.87	37.87	38.31	42.54
SD	11.80	9.76	11.13	10.16	10.16	10.14	9.05*
IR	4.56	2.38	1.88	3.73	3.73	3.78	**4.70**

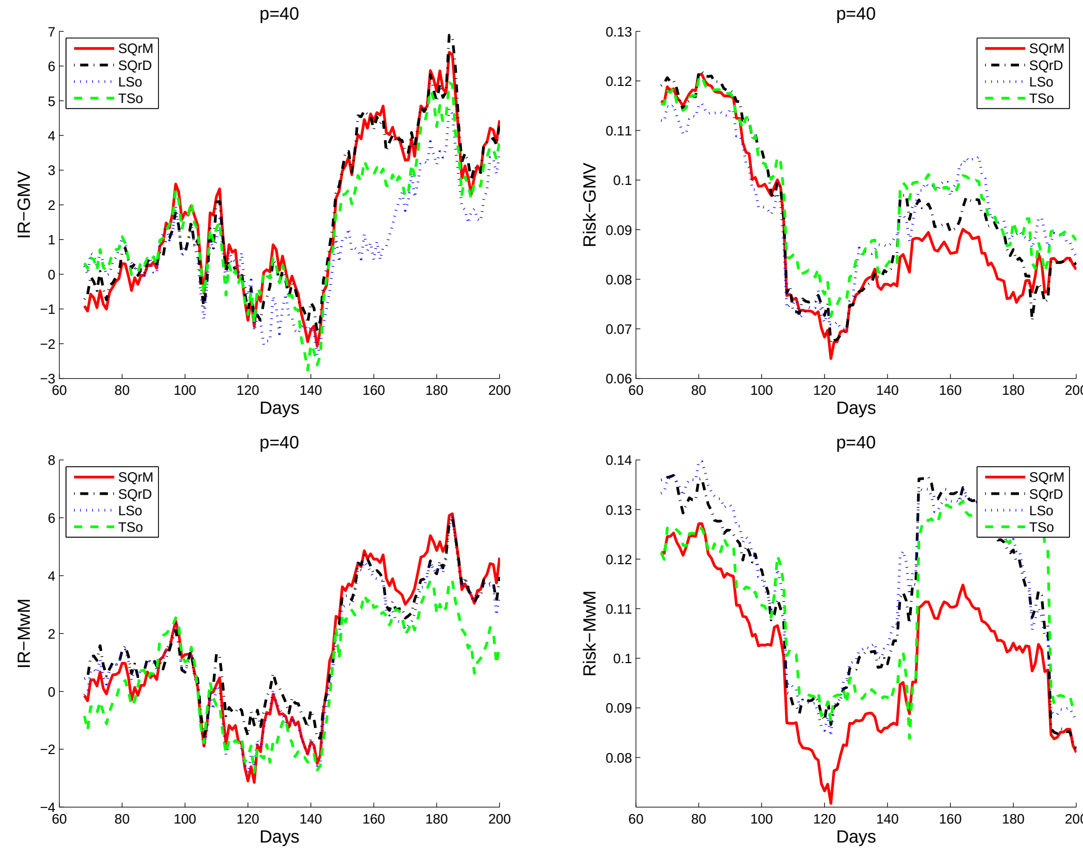
Note: AV, SD, IR denote the average, standard deviation, and information ratio of 87 daily log-returns respectively. AV, SD are annualized and in percent. The smallest numbers in the row labeled by SD is reported in bold face. SPo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the sample covariance matrix of daily log-returns (190, 230 and 130 days when $p = 30, 40$ and 50, respectively). LSo corresponds to the case where $\mathbf{\Sigma}_{T-h,T}$ is estimated by the linear shrinkage of the sample covariance matrix of daily log-returns (250 days when $p = 30, 40$ and 50).

Figure 1: Information ratios and standard deviations of log-returns of four strategies based on rolling windows of historical data for the GMV and MwM portfolios when $p = 30$.



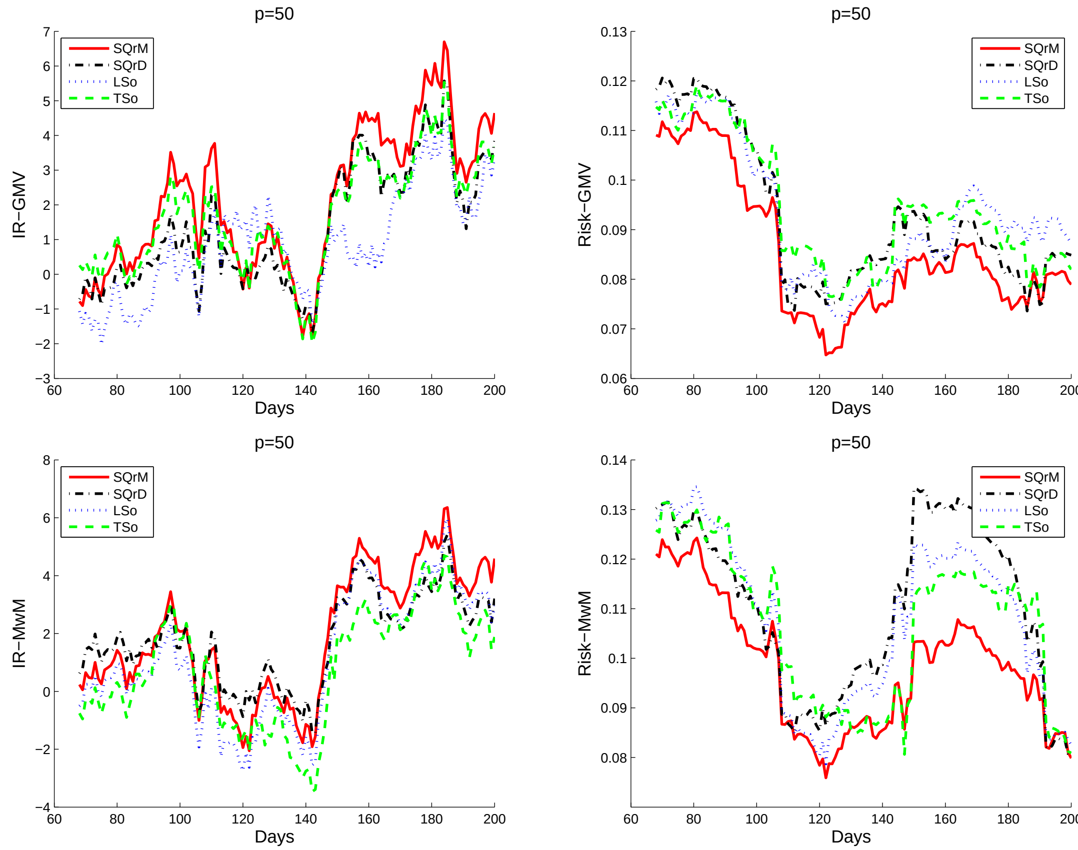
Note: Rolling windows of annualized standard deviations and information ratios of log-returns for the GMV and MwM portfolios. Each point is the standard deviation or information ratio of 42 log-returns of each portfolio strategy. Move one trading day forward at one time such that there are 133 different investment periods and each period contains 42 days (two months). The upper plots correspond to: SQrM uses 1 day of all intra-day data and 9 days of 15-minute data; SQrD uses 5 days of all intra-day data and 110 days of daily data; LSo uses 250 daily data; TSo uses 10 days of intra-day data. The bottom plots correspond to: SQrM use 5 days of all intra-day data and 14 days of 15-minute data; SQrD uses 2 days of all intra-day data and 90 days of daily data; LSo and SPo use 110 days of daily data.

Figure 2: Information ratios and standard deviations of log-returns of four strategies based on rolling windows of historical data for the GMV and MwM portfolios when $p = 40$.



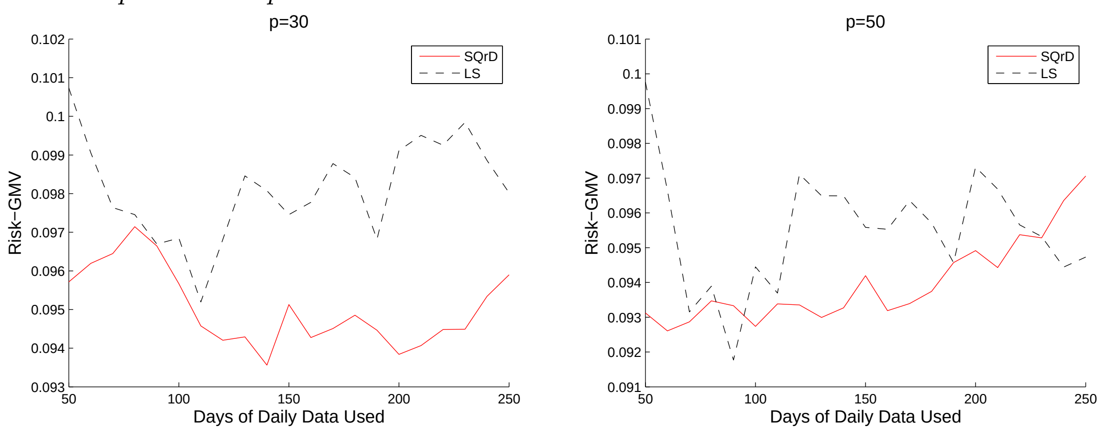
Note: Rolling windows of annualized standard deviations and information ratios of log-returns for the GMV and MwM portfolios. Each point is the standard deviation or information ratio of 42 log-returns of each portfolio strategy. Move one trading day forward at one time such that there are 133 different investment periods and each period contains 42 days (two months). The upper plots correspond to: SQrM uses 1 day of all intra-day data and 17 days of 15-minute data; SQrD uses 1 days of all intra-day data and 110 days of daily data; LSo uses 250 daily data; TSo uses 10 days of intra-day data. The bottom plots correspond to: SQrM use 4 days of all intra-day data and 15 days of 15-minute data; SQrD uses 5 days of all intra-day data and 200 days of daily data; LSo and SPo use 250 and 190 days of daily data, respectively.

Figure 3: Information ratios and standard deviations of log-returns of four strategies based on rolling window of historical data for the GMV and MwM portfolios when $p = 50$.



Note: Rolling windows of annualized standard deviations and information ratios of log-returns for the GMV and MwM portfolios. Each point is the standard deviation or the information ratio of 42 log-returns of each portfolio strategy. Move one trading day forward at one time such that there are 133 different investment periods and each period contains 42 days (two months). The upper plots correspond to: SQrM uses 1 day of all intra-day data and 15 days of 15-minute data; SQrD uses 1 days of all intra-day data and 130 days of daily data; LSo uses 250 daily data; TSo uses 9 days of intra-day data. The bottom plots correspond to: SQrM use 4 days of all intra-day data and 13 days of 15-minute data; SQrD uses 5 days of all intra-day data and 90 days of daily data; LSo and SPo use 250 and 130 days of daily data, respectively.

Figure 4: Standard deviation of log-returns of SQrD and LS in 174 investment days for the GMV when $p = 30$ and $p = 50$.



Note: Each point is the standard deviation of 174 log-returns of each portfolio strategy.