A nonparametric eigenvalue-regularized integrated covariance matrix estimator for asset return data

A Nonparametric Eigenvalue-Regularized Integrated Covariance Matrix Estimator for Asset Return Data

Clifford Lam*and Phoenix Feng†

Department of Statistics, London School of Economics and Political Science

Abstract

In high-frequency data analysis, the extreme eigenvalues of a realized covariance matrix are biased when its dimension p is large relative to the sample size n. Furthermore, with non-synchronous trading and contamination of microstructure noise, we propose a nonparametrically eigenvalue-regularized integrated covariance matrix estimator (NERIVE) which does not assume specific structures for the underlying integrated covariance matrix. We show that NERIVE is positive definite in probability, with extreme eigenvalues shrunk nonlinearly under the high dimensional framework $p/n \to c > 0$. We also prove that in portfolio allocation, the minimum variance optimal weight vector constructed using NERIVE has maximum exposure and actual risk upper bounds of order $p^{-1/2}$. Incidentally, the same maximum exposure bound is also satisfied by the theoretical minimum variance portfolio weights. All these results hold true also under a jump-diffusion model for the log-price processes with jumps removed using the wavelet method proposed in Fan and Wang (2007). They are further extended to accommodate the existence of pervasive factors such as a market factor under the setting $p^{3/2}/n \to c > 0$. The practical performance of NERIVE is illustrated by comparing to the usual two-scale realized covariance matrix as well as some other nonparametric alternatives using different simulation settings and a real data set.

Key words and phrases. High frequency data; Microstructure noise; Non-synchronous trading; Integrated covariance matrix; Minimum variance portfolio; Nonlinear shrinkage.

JEL classification: C13, C14, C55, C58.

*Clifford Lam is Associate Professor, Department of Statistics, London School of Economics. Email: C.Lam2@lse.ac.uk
†Phoenix Feng is PhD student, Department of Statistics, London School of Economics. Email: H.Feng2@lse.ac.uk

1 Introduction

In modern day finance, the so called tick-by-tick data on the prices of financial assets are readily available together with huge volume of other financial data. Advanced computational power and efficient data storage facilities mean that these data are analyzed on a daily basis by various market makers and academic researchers. While the Markowitz portfolio theory (Markowitz, 1952) is originally proposed for a finite number of assets using inter-day price data, the now easily accessible intra-day high frequency price data for a large number assets gives rise to new possibilities for efficient portfolio allocation, on top of the apparent increase in sample size for returns and volatility matrix estimation.

Certainly, the associated challenges for using high frequency data have to be overcome at the same time. One main challenge comes from the well documented market microstructure noise in the recorded tick-by-tick price data (Aït-Sahalia et al., 2005, Asparouhova et al., 2013). Another challenge comes from the non-synchronous trading times when more than one assets are considered. In terms of integrated covariance estimation, Xiu (2010) suggested a maximum likelihood approach for consistent estimation under market microstructure noise. Aït-Sahalia et al. (2010) proposed a quasi-maximum likelihood approach for estimating the covariance between two assets, while Zhang (2011) proposed a two or multi-scale covariance estimator to remove the bias accumulated due to the microstructure noise in the usual realized covariance formula, at the same time overcoming the non-synchronous trading times problem by using previous-tick times (see Section 2 also). Other attempts to overcome these two challenges together include Barndorff-Nielsen et al. (2011a) and Griffin and Oomen (2011), to name but a few.

When there are more than one asset to manage, the integrated covariance matrix for the asset returns is an important input for risk management or portfolio allocation. A large number of assets requires an estimation of a large integrated covariance matrix. Even in the simplest case of independent and identically distributed random vectors, random matrix theory tells us that the sample covariance matrix will have severely biased extreme eigenvalues (see chapter 5.2 of Bai and Silverstein (2010) for instance). To give a simple demonstration of how serious the bias problem can be, suppose we have independent and identically distributed p-dimensional random vectors $\mathbf{X} = (x_1, \ldots, x_n)^{\mathrm{T}}$ with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_p$, where $\mathbf{I}_p$ is the $p \times p$ identity matrix. The Marčenko-Pastur Law (Marčenko and Pastur, 1967) states that the density function of the limiting spectrum of the sample covariance matrix $\mathbf{S} = n^{-1}\mathbf{X}\mathbf{X}^{\mathrm{T}}$ as $p, n \to \infty$ with $p/n \to c > 0$, is

$$p_c(x) = \begin{cases} \frac{1}{2\pi x c \sigma^2}\sqrt{(b-x)(x-a)}, & a \le x \le b; \\ 0, & \text{otherwise}, \end{cases}$$

where $a = \sigma^2(1-\sqrt{c})^2$, $b = \sigma^2(1+\sqrt{c})^2$. See Bai and Silverstein (2009) Section 3.1 also. With this, say

$p = 25$ and $n = 500$, i.e., p is just 5% of n, the largest and smallest eigenvalues are 50% larger and 40% smaller than the corresponding population ones (i.e., σ^2) respectively. It means that a seemingly small p is enough already for the sample covariance matrix to suffer from significant distortions for the extreme eigenvalues, creating instability. When $\mathbf{\Sigma} \neq \sigma^2 \mathbf{I}_p$ the distortion can potentially be more severe.

To ameliorate the bias issue above, researchers propose different methods to reduce the dimension of the estimation problem, which is of order p^2, where p is the number of assets. Wang and Zou (2010), Tao et al. (2013) and Kim et al. (2016) assume a sparsity condition (perhaps after removing a market factor) and use thresholding to regularize different integrated covariance matrix estimators based on previous-tick times. Tao et al. (2011) uses a thresholded estimator to find a factor model structure for the daily dynamics of the integrated covariance matrix. These methods reduce the effective number of parameters to estimate to the order of p or less (or $p\log(p)$ for approximate sparsity. See Tao et al. (2013)). While consistent results are established for these methods, sparsity or factor model structure imposed regularities in the integrated covariance matrix which may not be completely satisfied in practice.

At the same time, with respect to portfolio allocation, DeMiguel et al. (2009) constrains the portfolio norm of a portfolio $\mathbf{w}$ using either the L_1 or squared L_2 norm, defined respectively by $\left\|\mathbf{w}\right\|_1 = \sum_i |w_i|$ and $\left\|\mathbf{w}\right\|_2^2 = \sum_i w_i^2$. Fan et al. (2012) proposes to regularize the portfolio weights by constraining the L_1 norm of the portfolio, termed the gross exposure of a portfolio in the paper. These two portfolio allocation methods do not regularize the integrated covariance matrix, but directly regularize the portfolio weights. The two-scale covariance matrix constructed in Fan et al. (2012) using the pairwise refresh method, however, may not be positive definite and adjustments are necessary to make it so again. In a very broad sense, these two methods are variations of sparsity or factor model-assumed papers mentioned in the previous paragraph, essentially reducing an order p^2 problem to order p or less by assuming a sparse optimal portfolio weight.

In this paper, we address the estimation of the integrated covariance matrix by reducing it to exactly an order p problem, but without assuming inherent structures to the population integrated covariance matrix or optimal portfolio weight. While this makes it impossible to estimate the integrated covariance matrix consistently, we achieve another important objective - regularization of extreme eigenvalues of the realized covariance matrix under the setting $p/n \to c > 0$ - through introducing a class of rotation-equivariant estimators and bringing it as close to the population counterpart as possible. Indeed, it is clear in our simulations and portfolio allocation exercises in Section 5 that the two-scale covariance matrix, which is essentially a realized covariance matrix, suffers from bad performance because of the instability created by the biases in its extreme eigenvalues compared to its population counterpart.

The said regularization above is achieved by minimizing a certain Frobenius error, to be discussed in Section 2.2. Such a regularization is inspired by a data splitting method originated from Abadir et al.

(2014), which is proved in Lam (2016) to be nonlinearly shrinking the sample covariance eigenvalues at a certain data splitting ratio. We show that the resulting integrated covariance matrix estimator is consistent with a certain positive definite matrix with regularized eigenvalues at a rate of $n^{-1/6}$ under the setting $p/n \to c > 0$, with n being the sample size. This is the same rate as the univariate two-scale realized covariance estimator by Zhang (2011). We also prove the same rate of convergence when there are pervasive factors but with $p^{3/2}/n \to c > 0$. Using its inverse in the construction of the minimum variance portfolio induces a natural upper bound on the maximum exposure of the portfolio, which decays at a rate of $p^{-1/2}$ in probability when there are no pervasive factors. The importance of this bound is that the theoretical minimum variance portfolio satisfies such a bound also. See Theorem 5 for more details, which includes results when there are pervasive factors like a market factor in the data.

The rest of the paper is organized as follows. Section 2 presents the notations and model for the high frequency data and introduces our way to perform nonlinear shrinkage on the two-scale covariance matrix estimator. Asymptotic theories and detailed assumptions, including those involving jumps removed data in the case of jump-diffusion log-price processes, can be found in Section 3. Practical concerns and implementation can be found in Section 4, while all simulations and a thorough empirical study are presented in Section 5. We give the conclusion of the paper in Section 6, before all the proofs of the theorems in the paper in Section 7.

2 Framework and Methodology

Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0\le t\le 1}, \mathbb{P}\right)$ be a filtered probability space on which the log-price process of the p assets under study, $\{\mathbf{X}_t\}_{0\le t\le 1}$, is adapted, where $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(p)})^{\mathrm{T}}$. We assume $\mathbf{X}_t$ follows a diffusion process

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

so that the time period is normalized to have length 1. Let L be the number of partitions of the data, with

$$0 = \tau_0 < \tau_1 < \cdots < \tau_L = 1,$$

and $(\tau_{\ell-1}, \tau_\ell]$ represents the ℓth partition. The reason we partition the data is that our method of regularization is carried out within a partition at a time, with data from outside of the partition help regularize the estimator within. The ultimate estimator is then the sum of all regularized estimators for the partitions. See section 2.2 for full details.

We assume that L is finite throughout the paper. The process $\{\mathbf{W}_t\}$ is a p-dimensional standard

Brownian motion. The drift $\boldsymbol{\mu}_t \in \mathbb{R}^p$ is càdlàg. It can be random and can be correlated with $\{\mathbf{W}_t\}$. The volatility $\boldsymbol{\sigma}_t \in \mathbb{R}^{p\times p}$ is also càdlàg. See all assumptions in Section 3 for full details. For each time interval $[a, b] \subset [0, 1]$, the corresponding integrated covariance matrix is defined as

$$\boldsymbol{\Sigma}(a, b) = \int_a^b \boldsymbol{\sigma}_u \boldsymbol{\sigma}_u^{\mathrm{T}} du.$$

This matrix is an important input in risk assessment and in Markowitz portfolio allocation. If we have a portfolio $\mathbf{w}$ which stays constant over a period of time $[a, b]$, then the risk of the portfolio over this period of time can be expressed as

$$R^{1/2}(\mathbf{w}) = (\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}(a, b)\mathbf{w})^{1/2} = \Big(\int_a^b \mathbf{w}^{\mathrm{T}}\boldsymbol{\sigma}_t\boldsymbol{\sigma}_t^{\mathrm{T}}\mathbf{w}\, dt\Big)^{1/2}.$$

The integrand $\mathbf{w}^{\mathrm{T}}\boldsymbol{\sigma}_t\boldsymbol{\sigma}_t^{\mathrm{T}}\mathbf{w}$ can be considered an instantaneous squared-risk at time t for $\mathbf{w}$, and hence $R(\mathbf{w})$ is a measure of the total risk accumulated over the period $[a, b]$. At the same time, in Markowitz portfolio allocation for instance, $\boldsymbol{\Sigma}(a, b)^{-1}$ is required for the construction of the minimum variance portfolio (see Section 3.2 for more details).

Let $\{v_s\}, 1 \le s \le nL$ be the set of all-refresh times for the log-prices in $\mathbf{X}_t$, where $n(\ell)$ is the number of all-refresh times at partition ℓ, with $\ell = 1, \ldots, L$, and $n = L^{-1}\sum_{\ell=1}^{L} n(\ell)$ is the average number of all-refresh times in a partition, which has the same order as the total sample size nL since L is finite. An all-refresh time v_s is the time when all assets have been traded at least once from the last all-refresh time v_{s-1}. Let $t_s^j \in (v_{s-1}, v_s]$ be the sth previous-tick time for the jth asset, which is the last trading time before or at v_s. For non-synchronous trading, $t_s^{j_1} \neq t_s^{j_2}$ for $j_1 \neq j_2$ in general. Also, high-frequency prices are typically contaminated by microstructure noise, so that at the all-refresh time v_s, we only observe

$$\mathbf{Y}(s) = \mathbf{X}(s) + \boldsymbol{\epsilon}(s),\ s = 1, \ldots, nL, \tag{2.2}$$

where $\mathbf{X}(s) = (X^{(1)}_{t_s^1}, \ldots, X^{(p)}_{t_s^p})^{\mathrm{T}}$ and $\boldsymbol{\epsilon}(s) = (\epsilon^{(1)}_{t_s^1}, \ldots, \epsilon^{(p)}_{t_s^p})^{\mathrm{T}}$, and $\boldsymbol{\epsilon}(\cdot)$ can be dependent on $\mathbf{X}(\cdot)$ in general (see the assumptions in Section 3). The underlying microstructure noise process $\{\boldsymbol{\epsilon}_t\}_{0\le t\le 1}$ is assumed to be adapted to $\{\mathcal{F}_t\}_{0\le t\le 1}$, so that the observed price process $\{\mathbf{Y}_t\}_{0\le t\le 1}$ is also adapted.

2.1 Two-Scale Covariance Estimator

Contamination of microstructure noise in high-frequency data means that the usual realized covariance is heavily biased. Hence in Zhang (2011), a two-scale covariance estimator (TSCV) is introduced to remove this bias. In this paper, we use a slightly modified multivariate version of the two-scale covariance estimator,

also by Zhang (2011). For $\ell = 1, \ldots, L$, define

$$\begin{aligned}
\widehat{\langle \mathbf{Y}, \mathbf{Y}^{\mathrm{T}} \rangle}_{\ell} &= [\mathbf{Y}, \mathbf{Y}^{\mathrm{T}}]_{\ell}^{(K)} - \frac{|S^{\ell}(K)|_K}{|S^{\ell}(1)|} [\mathbf{Y}, \mathbf{Y}^{\mathrm{T}}]_{\ell}^{(1)}, \text{ with} \\
\left([\mathbf{Y}, \mathbf{Y}^{\mathrm{T}}]_{\ell}^{(m)}\right)_{i,j} &= [Y^{(i)}, Y^{(j)}]_{\ell}^{(m)} = \frac{1}{m} \sum_{r \in S^{\ell}(m)} (Y_{t_r^i}^{(i)} - Y_{t_{r-m}^i}^{(i)})(Y_{t_r^j}^{(j)} - Y_{t_{r-m}^j}^{(j)}), \text{ and} \\
S^{\ell}(m) &= \{r : t_r^i, t_{r-m}^i \in (\tau_{\ell-1}, \tau_{\ell}] \text{ for all } i\}, \; |S^{\ell}(m)|_m = \frac{|S^{\ell}(m)| - m + 1}{m}.
\end{aligned} \tag{2.3}$$

Note that $[Y^{(i)}, Y^{(j)}]_{\ell}^{(1)}$ is the usual realized covariance matrix when returns are calculated using adjacent previous-tick times, whereas $[Y^{(i)}, Y^{(j)}]_{\ell}^{(K)}$ can be seen as a realized covariance matrix when returns are calculated at time points which are K previous-tick times apart instead of 1 (so, another scale). Ultimately, while both are dominated by the market microstructure noise, the difference defined in $\widehat{\langle \mathbf{Y}, \mathbf{Y}^{\mathrm{T}} \rangle}_{\ell}$ is proved in Zhang (2011) to be able to cancel out the dominating effect of the microstructure noise. With this, we define the two-scale covariance matrix (TSCV) for the partition $(\tau_{\ell-1}, \tau_{\ell}]$ to be

$$\widetilde{\mathbf{\Sigma}}(\tau_{\ell-1}, \tau_{\ell}) = \widehat{\langle \mathbf{Y}, \mathbf{Y}^{\mathrm{T}} \rangle}_{\ell}. \tag{2.4}$$

We suppress the dependence on K in the notation $\widetilde{\mathbf{\Sigma}}(\tau_{\ell-1}, \tau_{\ell})$ and all related definitions in the next section. In Section 3, we show that K works well at the order $n^{2/3}$, which is indeed the order of magnitude suggested in Zhang (2011).

Remark 1. *The multi-scale realized volatility matrix (MSRVM) by Tao et al. (2013), the kernel realized volatility matrix (KRVM) by Barndorff-Nielsen et al. (2011b) and the pre-averaging realized volatility matrix (PRVM) by Christensen et al. (2010) all have better convergence rates than the TSCV for multivariate settings. The latter two estimators can be constructed to be positive semi-definite, although all three estimators do not allow p to be growing with n. In principle, our regularized estimator, to be introduced in Section 2.2, can be based on regularizing these three estimators. However, while the proof of our regularization method on the MSRVM is an extension of ours on the TSCV (because MSRVM involves sums of order of $n^{1/2}$ terms), the jittering and pre-averaging operations on the KRVM and PRVM respectively are more difficult to handle in the proofs. We decide to leave the extensions of our regularization method to these estimators in a future project.*

2.2 Our Proposed Integrated Covariance Matrix Estimator

Although the two-scale covariance estimator in (2.4) removes the bias contributed from the microstructure noise, it does not solve the bias issue for the extreme eigenvalues when p is large such that $p/n \to c > 0$,

where the spread of the eigenvalues in the realized covariance matrix $\widetilde{\mathbf{\Sigma}}(\tau_{j-1}, \tau_j)$ is much larger than the population counterpart, creating instability. In Abadir et al. (2014), in a setting with stationary covariance matrix, they introduce the idea of splitting the data into two parts in order to regularize the sample covariance matrix constructed from one part of the data. Lam (2016) shows that with a certain splitting ratio, in fact the extreme eigenvalues of the sample covariance matrix are nonlinear shrunk asymptotically, the same as the nonlinear shrinkage introduced in Ledoit and Wolf (2012). We employ the data splitting idea in Abadir et al. (2014) for our high-frequency data setting in this paper. In order to regularize the realized covariance matrix in the time period $(\tau_{j-1}, \tau_j]$, $j = 1, \ldots, L$, we follow Lam (2016) and consider a rotation-equivariant estimator $\mathbf{\Sigma}(\mathbf{D}) = \mathbf{P}_{-j}\mathbf{D}\mathbf{P}_{-j}^{\mathrm{T}}$, where $\mathbf{D}$ is a diagonal matrix, and $\mathbf{P}_{-j}$ is orthogonal such that

$$\widetilde{\mathbf{\Sigma}}_{-j} = \mathbf{P}_{-j}\mathbf{D}_{-j}\mathbf{P}_{-j}^{\mathrm{T}}, \quad j = 1, \ldots, L, \quad \text{with } \widetilde{\mathbf{\Sigma}}_{-j} = \sum_{\ell \neq j} \widetilde{\mathbf{\Sigma}}(\tau_{\ell-1}, \tau_\ell). \tag{2.5}$$

The class of rotation-equivariant estimators allows for the same rotation of the estimator when the observed vectors are rotated. This is first introduced in James and Stein (1961) for estimating a covariance matrix under the Stein's loss function, with respect to which this class is invariant under rotation. Hence with no *a priori* information of the eigenvectors of the population covariance matrix, this class provides a good starting point as an estimator. Ledoit and Wolf (2012) used this class of estimators for the purpose of nonlinear shrinkage of eigenvalues. However, high frequency data vectors are in general not independent and identically distributed, so that the explicit nonlinear shrinkage formula in Ledoit and Wolf (2012) cannot be used.

To introduce our estimator, consider the following optimization problem, with similar problem considered in Ledoit and Wolf (2012) and Lam (2016):

$$\min_{\mathbf{D} \text{ diagonal}} \left\| \mathbf{P}_{-j}\mathbf{D}\mathbf{P}_{-j}^{\mathrm{T}} - \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \right\|_F, \tag{2.6}$$

where $\left\| \cdot \right\|_F$ denotes the Frobenius norm. Unlike Ledoit and Wolf (2012) which uses the eigenmatrix of the sample covariance constructed from the full data set, we use $\mathbf{P}_{-j}$ for the rotation-equivariant class. This facilitates regularization by allowing us to condition on the information outside of partition j, which weaken the correlation between $\{\mathbf{X}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$, and the serial correlation in $\{\boldsymbol{\epsilon}_t\}$ within partition j. See Assumption (E3) in Section 3.

Proposition 1. *The optimization problem (2.6) has solution* $\mathbf{D} = \mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1}, \tau_j)\mathbf{P}_{-j})$, *where* $\mathrm{diag}(A)$ *creates a diagonal matrix using the diagonal elements of* A.

Proof of Proposition 1. To simplify notations in this proof, write $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_p)$, $\mathbf{P}_{-j} = \mathbf{Q} =$

$(\mathbf{q}_1, \ldots, \mathbf{q}_p)$ and $\mathbf{\Sigma}(\tau_{j-1}, \tau_j) = \mathbf{\Sigma}_j$. Then

$$\left\|\mathbf{P}_{-j}\mathbf{D}\mathbf{P}_{-j}^{\mathrm{T}} - \mathbf{\Sigma}(\tau_{j-1}, \tau_j)\right\|_F^2 = \mathrm{tr}(\mathbf{D} - \mathbf{Q}^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{Q})^2 = \sum_{i=1}^{p} d_i^2 - 2\mathrm{tr}(\mathbf{D}\mathbf{Q}^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{Q}) + \mathrm{tr}(\mathbf{Q}^{\mathrm{T}}\mathbf{\Sigma}_j^2\mathbf{Q})$$
$$= \sum_{i=1}^{p} d_i^2 - 2\sum_{i=1}^{p} d_i\mathbf{q}_i^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{q}_i + \mathrm{tr}(\mathbf{Q}^{\mathrm{T}}\mathbf{\Sigma}_j^2\mathbf{Q}).$$

Differentiating the above with respect to d_i and set the derivative to 0, we get $d_i = \mathbf{q}_i^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{q}_i$, which leads to the solution $\mathbf{D} = \mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1}, \tau_j)\mathbf{P}_{-j})$. □

Clearly, all eigenvalues of $\mathbf{D}$ are contained within the largest and smallest eigenvalues of $\mathbf{\Sigma}(\tau_{j-1}, \tau_j)$. This way, the spread of the eigenvalues in $\mathbf{D}$ is regularized. Ultimately, we can prove that all the elements in $\mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\widetilde{\mathbf{\Sigma}}(\tau_{j-1}, \tau_j)\mathbf{P}_{-j})$ are asymptotically close to those in $\mathbf{D} = \mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1}, \tau_j)\mathbf{P}_{-j})$ in probability. See Theorem 2. This allows us to define our integrated covariance matrix estimator for the partition $(\tau_{j-1}, \tau_j]$ to be

$$\widehat{\mathbf{\Sigma}}(\tau_{j-1}, \tau_j) = \mathbf{P}_{-j}\mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\widetilde{\mathbf{\Sigma}}(\tau_{j-1}, \tau_j)\mathbf{P}_{-j})\mathbf{P}_{-j}^{\mathrm{T}}. \tag{2.7}$$

The overall integrated covariance matrix estimator for the period $[0, 1]$ is then defined to be

$$\widehat{\mathbf{\Sigma}}(0, 1) = \sum_{j=1}^{L} \widehat{\mathbf{\Sigma}}(\tau_{j-1}, \tau_j) = \sum_{j=1}^{L} \mathbf{P}_{-j}\mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\widetilde{\mathbf{\Sigma}}(\tau_{j-1}, \tau_j)\mathbf{P}_{-j})\mathbf{P}_{-j}^{\mathrm{T}}. \tag{2.8}$$

An ideal estimator relative to $\widehat{\mathbf{\Sigma}}(0, 1)$ is an estimator with $\widetilde{\mathbf{\Sigma}}(\tau_{j-1}, \tau_j)$ replaced by the population counterpart $\mathbf{\Sigma}(\tau_{j-1}, \tau_j)$, i.e.,

$$\mathbf{\Sigma}_{\mathrm{Ideal}}(0, 1) = \sum_{j=1}^{L} \mathbf{P}_{-j}\mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1}, \tau_j)\mathbf{P}_{-j})\mathbf{P}_{-j}^{\mathrm{T}}. \tag{2.9}$$

In practice, a partition can be a trading day or a quarter of it, depending on the number of trading days of data we have and the number of all-refresh data points in them. We proposed an optimization criterion to choose the number of partitions (not necessarily uniform) in Section 4. In our simulations and empirical examples in Section 5, we use 5 or 1 day of training data with $(\tau_{\ell-1}, \tau_\ell]$ set at 1 day or a quarter of a day, with the number of all-refresh data points in the order of hundreds in each interval.

3 Asymptotic Theory

In this section, we show that our proposed estimator (2.7) in the jth partition of the data is asymptotically close to the corresponding ideal rotation-equivariant estimator

$$\mathbf{\Sigma}_{\mathrm{Ideal}}(\tau_{j-1}, \tau_j) = \mathbf{P}_{-j}\mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1}, \tau_j)\mathbf{P}_{-j})\mathbf{P}_{-j}^{\mathrm{T}}. \tag{3.1}$$

This is exactly the optimal estimator that solves (2.6) following Proposition 1. While $\widetilde{\mathbf{\Sigma}}(\tau_{j-1},\tau_j)$ can have its spread of eigenvalues much larger than that of $\mathbf{\Sigma}(\tau_{j-1},\tau_j)$ when $p/n \to c > 0$, our estimator $\widehat{\mathbf{\Sigma}}(\tau_{j-1},\tau_j)$ in (2.7) has its spread of eigenvalues contained within the spread of $\mathbf{\Sigma}(\tau_{j-1},\tau_j)$ asymptotically by being close to $\mathbf{\Sigma}_{\text{Ideal}}(\tau_{j-1},\tau_j)$ in (3.1) above. See Theorem 2 below. We first introduce some assumptions for our theorems to hold. In the following and hereafter, we denote $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ the minimum and maximum eigenvalue of a square matrix respectively. We write $a \asymp b$ to mean that $a = O(b)$ and $b = O(a)$, and $a \asymp_P b$ to mean that $a = O_P(b)$ and $b = O_P(a)$.

For $j = 1, \ldots, L$, and $v_s = v_s^j$ which is the sth all-refresh time within partition j, define

$$\mathcal{F}_{-j} = \mathcal{F}_{\tau_{j-1}} \cup \mathcal{F}/\mathcal{F}_{\tau_j}, \quad \mathcal{F}_s^j = \mathcal{F}_{v_s}/\mathcal{F}_{\tau_{j-1}},$$

with $\mathcal{F}_s^j = \phi$ for $s \leq 0$. The following assumptions are true for $K = 1$ or $K \asymp n^{2/3}$.

Assumptions on the drift $\boldsymbol{\mu}_t$:

(D1) The drift $\boldsymbol{\mu}_t$ has càdlàg components, such that for $s = K, K+1, \ldots, n(j)$,

$$\int_{v_{s-K}}^{v_s} \boldsymbol{\mu}_t dt = \mathbf{A}(v_{s-K}, v_s)\mathbf{Z}_{d,s}^j,$$

where $\mathbf{A}(v_{s-K}, v_s) \neq \mathbf{0}$ is a non-random $p \times p$ matrix and can be asymmetric and singular. It has $\big\|\mathbf{A}(v_{s-K}, v_s)\big\| = O(p^{1/2}K^{1/2}|v_s - v_{s-1}|)$, where the order $p^{1/2}$ only appears when there are only finite number of columns (say r) that are non-zero. The random vector $\mathbf{Z}_{d,s}^j \in \mathcal{F}_s^j$ has components conditionally independent of each other given $\mathcal{F}_{-j}$, with eighth moments exist. Also, $E(\mathbf{Z}_{d,s}^j|\mathcal{F}_{-j}) = \mathbf{0}$ and $\text{var}(\mathbf{Z}_{d,s}^j|\mathcal{F}_{-j}) = \mathbf{I}_p$ almost surely.

The drift $\boldsymbol{\mu}_t$ can also be non-random, in which case $\mathbf{Z}_{d,s}^j = (1, 0, \ldots, 0)^{\mathrm{T}}$ for all s, and the assumption for $\mathbf{A}(v_{s-K}, v_s)$ is the same as above.

(D2) Write $\mathbf{P}_{-j} = (\mathbf{p}_{1j}, \ldots, \mathbf{p}_{pj})$. We assume for each $i = 1, \ldots, p$, and $s = rK + q$ for $r = 1, \ldots, |S^j(K)|_K$ and $q = 0, 1, \ldots, K-1$, there exists $\rho_{d,K,q}^j \in \mathcal{F}_{-j}$ such that $0 \leq \rho_{d,K,q}^j \leq \xi < 1$ with ξ a constant, and for $\ell = K + q, 2K + q, \ldots, rK + q$,

$$\begin{aligned}
&E\big((\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K}, v_s)\mathbf{Z}_{d,\ell}^j)^2|\mathcal{F}_{-j} \cup \mathcal{F}_{\ell-K}^j\big) \\
&= \rho_{d,K,q}^j(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K}, v_s)\mathbf{Z}_{d,\ell-K}^j)^2 + (1 - \rho_{d,K,q}^j)\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K}, v_s)\mathbf{A}(v_{s-K}, v_s)^{\mathrm{T}}\mathbf{p}_{ij} + e_{d,\ell-K}^{ij},
\end{aligned}$$

where we define $\mathbf{Z}_{d,\ell}^j\mathbf{Z}_{d,\ell}^{j\mathrm{T}} = \mathbf{I}_p$ and $e_{d,\ell}^{ij} = 0$ for $\ell \leq 0$. The process $\{e_{d,\ell}^{ij}\}$ with $e_{d,\ell}^{ij} \in \mathcal{F}_\ell^j$ has $E(e_{d,\ell}^{ij}|\mathcal{F}_{-j} \cup \mathcal{F}_{\ell-K}^j) = 0$ almost surely, and $e_{d,\ell}^{ij}|\mathcal{F}_{-j} \cup \mathcal{F}_{\ell-K}^j = O_P(\big\|\mathbf{A}(v_{s-K}, v_s)\big\|^2)$.

(D3) Let $\psi(x) = e^{x^2} - 1$. We assume that for $\ell = 0, 1, \ldots, s$,

$$E\left\{\psi\left(\frac{|(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K}, v_s)\mathbf{Z}_{d,\ell}^j)^2 - \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K}, v_s)\mathbf{A}(v_{s-K}, v_s)^{\mathrm{T}}\mathbf{p}_{ij}|}{(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K}, v_s)\mathbf{Z}_{d,\ell-K}^j)^2}\right)\Big|\mathcal{F}_{-j} \cup \mathcal{F}_{\ell-K}^j\right\} < \infty,$$
$$E\left\{\psi\left(\frac{|e_{d,\ell}^{ij}|}{(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K}, v_s)\mathbf{Z}_{d,\ell-K}^j)^2}\right)\Big|\mathcal{F}_{-j} \cup \mathcal{F}_{\ell-K}^j\right\} < \infty.$$

Assumptions on the volatility $\boldsymbol{\sigma}_t$ and Brownian motion $\mathbf{W}_t$:

(V1) The volatility $\boldsymbol{\sigma}_t$ has càdlàg components, and the Brownian motion $\{\mathbf{W}_t\}$ can be correlated with $\{\boldsymbol{\mu}_t\}$ in general. Write

$$\int_{v_{s-K}}^{v_s} \boldsymbol{\sigma}_t d\mathbf{W}_t = \boldsymbol{\Sigma}(v_{s-K}, v_s)^{1/2}\mathbf{Z}_{v,s}^j,$$

where $\boldsymbol{\Sigma}(v_{s-K}, v_s)$ is a symmetric positive definite $p \times p$ matrix which can be random, with

$$\lambda_{\min}(\boldsymbol{\Sigma}(\tau_{j-1}, \tau_j)) \geq C(\tau_j - \tau_{j-1})^{-1}, \quad \lambda_{\max}(\boldsymbol{\Sigma}(v_{s-K}, v_s)) \asymp_P \left\|\mathbf{A}(v_{s-K}, v_s)\right\|^2/|v_s - v_{s-K}|,$$

where $C > 0$ is a constant. The process $\{\boldsymbol{\sigma}_t\}$ is independent of all other processes.

Also, $E(\mathbf{Z}_{v,s}^j|\mathcal{F}_{-j}) = \mathbf{0}$ and $\mathrm{var}(\mathbf{Z}_{v,s}^j|\mathcal{F}_{-j}) = \mathbf{I}_p$ almost surely. The random vector $\mathbf{Z}_{v,s}^j \in \mathcal{F}_s^j$ has components conditionally independent of each other given $\mathcal{F}_{-j}$, with eighth moments exist.

(V2) Parallel to (D2), but expectations are taken conditional on $\mathcal{F}_{-j} \cup \mathcal{F}_{\ell-K}^j \cup \mathcal{F}_{v_s}^{\boldsymbol{\sigma}}$, where $\mathcal{F}_t^{\boldsymbol{\sigma}}$ is the σ-algebra generated by the process $\{\boldsymbol{\sigma}_t\}$ up to time t.

Also, $\rho_{d,K,q}^j$ is replaced by $\rho_{v,K,q}^j \in \mathcal{F}_{-j}$, $\mathbf{A}(v_{s-K}, v_s)$ by $\boldsymbol{\Sigma}(v_{s-K}, v_s)^{1/2}$, $\mathbf{Z}_{d,\ell}^j$ by $\mathbf{Z}_{v,\ell}^j$ and $e_{d,\ell}^{ij}$ by $e_{v,\ell}^{ij}$ with $e_{v,\ell}^{ij}|\mathcal{F}_{-j} \cup \mathcal{F}_{\ell-K}^j \cup \mathcal{F}_{v_s}^{\boldsymbol{\sigma}} = O_P(e_{d,\ell}^{ij}|\mathcal{F}_{-j} \cup \mathcal{F}_{\ell-K}^j)/|v_s - v_{s-K}|$.

(V3) Parallel to (D3), replacements the same as in (V2).

Assumptions on the microstructure noise $\boldsymbol{\epsilon}_t$:

(E1) Within the jth partition, $E(\boldsymbol{\epsilon}(s)\boldsymbol{\epsilon}(s)^{\mathrm{T}}|\mathcal{F}_{-j}) = \boldsymbol{\Sigma}_{\epsilon,s}^j$, which is random and independent of all other processes given $\mathcal{F}_{-j}$. Also, $E(\boldsymbol{\Sigma}_{\epsilon,s}^j) = \boldsymbol{\Sigma}_\epsilon^j$, and $\left\|\boldsymbol{\Sigma}_{\epsilon,s}^j\right\| \leq \lambda_\epsilon < \infty$ uniformly as $n, p \to \infty$ where λ_ϵ is a constant.

(E2) Within the jth partition, we can write $\boldsymbol{\epsilon}(s) = (\boldsymbol{\Sigma}_{\epsilon,s}^j)^{1/2}\mathbf{Z}_{\epsilon,s}^j$, with $\mathbf{Z}_{\epsilon,s}^j \in \mathcal{F}_s^j$ having conditionally independent components given $\mathcal{F}_{-j}$. Also $E(\mathbf{Z}_{\epsilon,s}^j|\mathcal{F}_{-j}) = 0$ almost surely and eighth order moments exist for the components of $\mathbf{Z}_{\epsilon,s}^j$.

(E3) Let $\mathcal{F}_t^X$ be the σ-algebra generated by the log-price process up to time t, and $\mathcal{F}_t^\epsilon$ the one by the microstructure noise process up to time t, so that $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s^X \otimes \mathcal{F}_s^\epsilon$. Then for s_1, s_2 time points within partition j, given $\mathcal{F}_{-j}$, we assume the φ-mixing coefficient between two σ-algebras satisfies

$$\varphi(\mathcal{F}_{s_1}^X, \mathcal{F}_{s_2}^\epsilon | \mathcal{F}_{-j}) = O(n^{-1}) = \varphi(\mathcal{F}_{s_2}^\epsilon, \mathcal{F}_{s_1}^X | \mathcal{F}_{-j}).$$

Also, for $s_2 > s_1$ time points within partition j, we assume

$$\varphi(\mathcal{F}_{s_1}^\epsilon, \mathcal{F}_{s_2}^\epsilon / \mathcal{F}_{s_1}^\epsilon | \mathcal{F}_{-j}) = O(n^{-1}) = \varphi(\mathcal{F}_{s_2}^\epsilon / \mathcal{F}_{s_1}^\epsilon, \mathcal{F}_{s_1}^\epsilon | \mathcal{F}_{-j}).$$

Other assumptions:

(A1) The observation times are independent of $\mathbf{X}(\cdot)$ and $\boldsymbol{\epsilon}(\cdot)$, and the partition boundaries τ_ℓ, $\ell = 0, 1, \ldots, L$, satisfy $0 < C_1 \leq \min_{\ell=1,\ldots,L} L(\tau_\ell - \tau_{\ell-1}) \leq \max_{\ell=1,\ldots,L} L(\tau_\ell - \tau_{\ell-1}) \leq C_2 < \infty$, where C_1, C_2 are generic constants. Also, the all-refresh times v_s, $s = 1, \ldots, nL$ satisfy $\max_{s=1,\ldots,nL} nL(v_s - v_{s-1}) \leq C_3$ for a generic constant $C_3 > 0$. Moreover, $\max_{\ell=1,\ldots,L} nL(\tau_\ell - v_{n(\ell)}) = o(1)$. The sample size in the jth partition has $n(j)/n \to 1$.

(A2) The pervasive factors, if any, persist within an interval $(v_{s-1}, v_s]$ for $s = 1, \ldots, nL$.

There is another set of assumptions (A3) to (A5) in Section 7. They involve the drift and volatility in $\mathbf{X}_{v_s} - \mathbf{X}(s)$, i.e., the drift and volatility in between the all-refresh and the previous-tick times. These assumptions are in many ways parallel to assumptions (D1) to (D3) and (V1) to (V3), but the decompositions are more involved, so that we choose to present them in Section 7 to aid the flow of the paper.

The matrix $\mathbf{A}(v_{s-K}, v_s)$ in assumptions (D1) to (D3) plays the role of a factor loading matrix in a factor model if the drift $\boldsymbol{\mu}_t$ is random. Within partition j, if $\mathbf{A}(v_{s-K}, v_s)$ is diagonal, the contribution of drift among all assets over v_{s-K} to v_s are conditionally independent given $\mathcal{F}_{-j}$. If $\mathbf{A}(v_{s-K}, v_s)$ is singular with only the first $r \ll p$ columns being non-zero, then it represents an exact r-factor model with no noise on the drift. The first r singular values of $\mathbf{A}(v_{s-K}, v_s)$ are then of order $p^{1/2}K^{1/2}|v_s - v_{s-1}|$, with $K^{1/2}|v_s - v_{s-1}|$ accounting for the length of the time interval considered.

The serial dependence of the drift vector is depicted in Assumption (D2). This assumption is more general then it seems. For instance, $\mathbf{Z}_{d,\ell}^j$ can be a random vector of martingales, so that

$$E(\mathbf{Z}_{d,\ell}^j | \mathcal{F}_{-j} \cup \mathcal{F}_{\ell-K}^j) = \mathbf{Z}_{d,\ell-K}^j,$$

and hence $E(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{Z}_{d,\ell}^j|\mathcal{F}_{-j}\cup\mathcal{F}_{\ell-K}^j)=\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{Z}_{d,\ell-K}^j$. Then by Jensen's inequality,

$$E\big((\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{Z}_{d,\ell}^j)^2|\mathcal{F}_{-j}\cup\mathcal{F}_{\ell-K}^j\big)\geq(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{Z}_{d,\ell-K}^j)^2,$$

and the assumption only requires a uniformly strict inequality above, so that $\rho_{d,K,q}^j$ can be uniformly smaller than 1. Note also $E((\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{Z}_{d,\ell}^j)^2|\mathcal{F}_{-j})=\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{A}(v_{s-K},v_s)^{\mathrm{T}}\mathbf{p}_{ij}$, and hence the assumption balances this mean with the squared-conditional expected value of the martingale, subject to an error $e_{d,\ell}^{ij}$.

If $\mathbf{Z}_{d,\ell}^j$ is independent of any past information such that $E(\mathbf{Z}_{d,\ell}^j|\mathcal{F}_{-j}\cup\mathcal{F}_{\ell-K}^j)=\mathbf{0}$, then

$$E\big((\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{Z}_{d,\ell}^j)^2|\mathcal{F}_{-j}\cup\mathcal{F}_{\ell-K}^j\big)=\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{A}(v_{s-K},v_s)^{\mathrm{T}}\mathbf{p}_{ij},$$

so that Assumption (D2) means that $\rho_{d,K,q}^j=e_{d,\ell}^{ij}=0$.

Assumption (D3) says that quadratic forms not too far in time apart can be very different but with sub-Gaussian-tailed probability. Assumptions (D1) to (D3) together allow us to use certain Hoeffding's inequalities for sums of martingale differences (see van de Geer (2002), Theorem 2.2).

If the drift $\boldsymbol{\mu}_t$ is non-random, then the matrix $\mathbf{A}(v_{s-K},v_s)$ can be set as zero except the first column which is a non-zero known vector. With $\mathbf{Z}_{d,s}^j=(1,0,\ldots,0)^{\mathrm{T}}$, assumptions (D2) and (D3) are automatically satisfied with $e_{d,\ell}^{ij}=0$. We do not make further assumptions for $\mathbf{A}(\cdot,\cdot)$, and hence the drift can include longer term trends (where components of $\mathbf{A}(\cdot,\cdot)$ can be increasing or decreasing over different time segments) and pervasive factors.

Assumption (V1) to (V3) for the volatility are parallel to (D1) to (D3). The subtler part is in Assumption (V1), where $\big\|\boldsymbol{\Sigma}(v_{s-K},v_s)\big\|$ depends on $\big\|\mathbf{A}(v_{s-K},v_s)\big\|$. In doing so, we are essentially assuming that if there are pervasive factors such as the market factor, then they affect both the drift and the volatility of the log-price process at the same time, which certainly makes sense. Then order $p^{1/2}$ singular values in $\mathbf{A}(v_{s-K},v_s)$ translates to order p eigenvalues in $\boldsymbol{\Sigma}(v_{s-K},v_s)$ in the presence of pervasive factors, appropriately adjusted by $|v_s-v_{s-K}|$.

Assumption (E1) allows for time-varying covariance matrix for the microstructure noise. Assumption (E3) particularly assumes a weak dependence between the log-price process and the microstructure noise process within partition j, as well as a weak serial dependence among the microstructure noise vectors, when $\mathcal{F}_{-j}$ is given. This assumption is inspired by Chen and Mykland (2017), where they assumed that given the entire information of the log-price process, the microstructure noise at different time points are independent. In our case, we are not given the entire picture of the log-price process, but not far from that

either since with $\mathcal{F}_{-j}$ we are given $nL - n(j)$ data points from the total of nL. Then instead of assuming the microstructure noise vectors are independent, we assume that they are weakly dependent, and with n larger (i.e., more information at more time points is available outside partition j), the dependence is weaker.

The first part of Assumption (A1) is automatically satisfied if the boundary set $\{\tau_\ell\}_{0\le\ell\le L}$ is pre-set, for instance, to be the daily opening or closing time of the L days of data, or a quarter of it, just as described in Section 2.2. See also Section 4 on a criterion in choosing these tuning parameters. Assumption (A2) means that the pervasive factors are either present between two all-refresh times, or they are absent.

Theorem 2. *Let Assumptions (D1) to (D3), (V1) to (V3), (E1) to (E3) and (A1) to (A5) hold. For the all-refresh log-price data $\mathbf{Y}(s)$, $s = 1, \ldots, nL$ in (2.2), as $n, p \to \infty$ such that $p/n \to c > 0$, if there are no pervasive factors, i.e., $\left\|\mathbf{A}(v_{s-K}, v_s)\right\| = O(K^{1/2}|v_s - v_{s-1}|)$, the integrated covariance matrix estimator constructed in (2.7) and $\widehat{\mathbf{\Sigma}}(0,1)$ in (2.8) satisfy*

$$\max_{j=1,\ldots,L} \left\|\widehat{\mathbf{\Sigma}}(\tau_{j-1}, \tau_j)\mathbf{\Sigma}_{\text{Ideal}}(\tau_{j-1}, \tau_j)^{-1} - \mathbf{I}_p\right\| = O_P(n^{-1/6}),$$
$$\|\widehat{\mathbf{\Sigma}}(0,1)\mathbf{\Sigma}_{\text{Ideal}}(0,1)^{-1} - \mathbf{I}_p\| = O_P(n^{-1/6}),$$

where $\|\cdot\|$ denotes the spectral norm of a matrix. If there are pervasive factors so that $\left\|\mathbf{A}(v_{s-K}, v_s)\right\| = O(p^{1/2}K^{1/2}|v_s - v_{s-1}|)$ (this includes the case when $\boldsymbol{\mu}_t$ is assumed non-random), then assuming $p^{3/2}/n \to c > 0$, the above results still hold.

The proof can be found in Section 7. The rate of convergence of our estimator is $n^{-1/6}$, the same as the TSCV in the univariate case (Zhang, 2011). Note that Assumption (D1) and (V1) allow for the existence of pervasive factors like the market factor, and our estimator is still converging to the ideal estimator in probability at a rate of $n^{-1/6}$ if $p^{3/2}/n \to c > 0$. One remarkable fact is that this rate does not depend on p. We require p to be growing slower than n in the presence of pervasive factors mainly because the drift term can overwhelm the estimator when there are pervasive factors. When the drift is non-random under Assumption (D1), it certainly can behave as if there are pervasive factors when there are no further assumptions on $\mathbf{A}(\cdot,\cdot)$, and we do need $p^{3/2}/n \to c > 0$ for the results in Theorem 2 to hold. See Remark 3 also at the end of this Section as well. Indeed, without a drift term, Lam (2016) allows the (low frequency) data to have a factor structure under $p/n \to c > 0$.

Since $\mathbf{P}_{-j}$ is orthogonal, it is easy to see that $\mathbf{\Sigma}_{\text{Ideal}}(\tau_{j-1}, \tau_j)$ in (3.1) has

$$\text{Cond}(\mathbf{\Sigma}_{\text{Ideal}}(\tau_{j-1}, \tau_{j-1})) \le \text{Cond}(\mathbf{\Sigma}(\tau_{j-1}, \tau_j)),$$

where Cond$(\cdot)$ is the condition number of a matrix, defined by dividing the maximum over the minimum magnitude of eigenvalue of the matrix. Theorem 2 then implies that

$$\text{Cond}(\widehat{\mathbf{\Sigma}}(\tau_{j-1},\tau_{j-1})) \leq \text{Cond}(\mathbf{\Sigma}(\tau_{j-1},\tau_j))$$

in probability. This is the result of nonlinear shrinkage of the eigenvalues in $\widehat{\mathbf{\Sigma}}(\tau_{j-1},\tau_j)$. Our estimator then has its spread of eigenvalues contained within the population counterpart, so that it is more stable than $\widetilde{\mathbf{\Sigma}}(\tau_{j-1},\tau_j)$, which can have its extreme eigenvalues severely biased when $p/n \to c > 0$, creating instability. The TSCV indeed performs worse than all other methods in Section 5. Incidentally, since all eigenvalues of $\mathbf{\Sigma}(\tau_{j-1},\tau_j)$ are non-negative, the results of Theorem 2 also prove the following.

Corollary 3. *Let all the assumptions in Theorem 2 hold. Then as $n,p \to \infty$ such that $p/n \to c > 0$, the integrated covariance matrix estimator $\widehat{\mathbf{\Sigma}}(\tau_{j-1},\tau_j)$ in (2.7), and also $\widehat{\mathbf{\Sigma}}(0,1)$ in (2.8), are positive definite in probability as long as $\mathbf{\Sigma}(\tau_{j-1},\tau_j)$ and $\mathbf{\Sigma}(0,1)$ are.*

This corollary shows that the positive definiteness of an integrated covariance matrix is preserved in our proposed estimator in probability as we have large enough sample size. In practice, we always have positive definiteness of the estimator with a moderate sample size n and a similar dimension p.

Remark 2. *In Theorem 2, unlike Lam (2016), we do not require the partition to be very small with the number of data points of order smaller than the total sample size. This is because we are not proving efficiency relative to using the majority of data points in constructing the eigenmatrix for our rotation-equivariant estimator. We can pursue it, but then a very small partition essentially means $L \to \infty$ also, which unfortunately makes the rate of convergence to be slower than $n^{-1/6}$ due to the complications of microstructure noise. This can be seen explicitly in the proof of Lemma 4, where one of the term has rate $n^{-1/6}L$. The practical performance is also worse if we use a very small partition, resulting in too many of them. Hence we decide not to pursue something like Theorem 5 of Lam (2016), for the sake of a better rate of convergence, and a better practical performance overall.*

Remark 3. *Defining $\mathbf{p}_{ij}$ as an eigenvector for $\mathbf{P}_{-j}$, the term $\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{A}(v_{s-K},v_s)^{\mathrm{T}}\mathbf{p}_{ij}$ is bounded by $\left\|\mathbf{A}(v_{s-K},v_s)\right\|^2$ in our proofs when there are pervasive factors, which is an order p larger than when there are no factors. The same treatment goes when $\boldsymbol{\mu}_t$ is assumed non-random, where $\mathbf{A}(\cdot,\cdot)$ essentially has only one non-zero column. In the end, this is exactly the reason why $p^{3/2}/n \to c > 0$ is needed instead of just $p/n \to c > 0$. We conjecture that $p/n \to c > 0$ is enough for our results to hold even with pervasive factors, since $\mathbf{p}_{ij}$ is in fact a random eigenvector of a sample covariance-like matrix $\sum_{\ell \neq j}\widetilde{\mathbf{\Sigma}}(\tau_{\ell-1},\tau_\ell)$. If it were a proper sample covariance matrix, then for any known unit vector $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{x} = O_P(p^{-1/2})$ (see*

Theorem 1 and Remark 1 of Bai et al. (2007)), so that $\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-K},v_s)\mathbf{A}(v_{s-K},v_s)^{\mathrm{T}}\mathbf{p}_{ij}$ *should be of order* $\left\|\mathbf{A}(v_{s-K},v_s)\right\|^2/p$ *in probability, i.e., the same order as when there are no factors.*

3.1 Extension to Jump-Diffusion Processes

Our method can be extended to include jumps in the underlying log-price process $\mathbf{X}_t$. We introduce the relevant model first. With jumps, the underlying log-price process is modeled as

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\mathbf{W}_t + d\mathbf{J}_t, \quad t \in [0,1], \tag{3.2}$$

where $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ are as in the pure diffusion model (2.1), and $\mathbf{J}_t = (J_t^{(1)},\ldots,J_t^{(p)})^{\mathrm{T}}$ denotes a p-dimensional right-continuous pure jump process. Each element in $\mathbf{J}_t$ is assumed to have finite activity in $[0,1]$, so that there are only finite number of jumps in each log-price process $X_t^{(j)}$ in the time interval we consider. The $J_t^{(j)}$'s can be correlated with each other, and each is modeled by

$$J_t^{(j)} = \sum_{\ell=1}^{N_t^{(j)}} B_\ell^{(j)}, \quad t \in [0,1],$$

where each count process $N_t^{(j)}$ can be correlated with each other. The same holds true for each jump size $B_\ell^{(j)}$. The quadratic covariation over $[0,1]$ for the process $\mathbf{X}_t$ is then

$$QV = \int_0^1 \boldsymbol{\sigma}_t\boldsymbol{\sigma}_t^{\mathrm{T}} dt + \sum_{0\le t\le 1} \Delta\mathbf{J}_t\Delta\mathbf{J}_t^{\mathrm{T}}, \tag{3.3}$$

where $\Delta\mathbf{J}_t = \mathbf{J}_t - \mathbf{J}_{t-}$. It is clear that an off-diagonal entry in $\Delta\mathbf{J}_t\Delta\mathbf{J}_t^{\mathrm{T}}$ will only be non-zero in general when both the corresponding log-price processes have jumps at the same time (cojumps) for at least once. It can correspond to, e.g., certain major market news reacted by a number of stocks at the same time. To account for the jump risks contributed by regular occurrence of cojumps (see e.g. Gilder et al. (2014) for examples of systematic or non-systematic cojumps), QV should be estimated as a whole rather than just the integrated covariance matrix.

To this end, we propose to use the wavelet method described in Section 3.2 of Fan and Wang (2007) to first remove the jumps in the log-price processes and construct our nonlinear shrinkage estimator in (2.8) using the jumps-removed data. The wavelet approach is also considered in Xue et al. (2014) to test for the presence of jumps in high-frequency financial time series. We give the practical details on how we implement the wavelet method for each observed log-price process at the end of the section. The estimated jump process $\widehat{\mathbf{J}}_t$ using the wavelet method is then used to construct $\sum_{0\le t\le 1}\Delta\widehat{\mathbf{J}}_t\Delta\widehat{\mathbf{J}}_t^{\mathrm{T}}$, giving us

an estimator of QV as

$$\widehat{QV} = \widehat{\mathbf{\Sigma}}(0,1) + \sum_{0\le t\le 1} \Delta\widehat{\mathbf{J}}_t \Delta\widehat{\mathbf{J}}_t^{\mathrm{T}}. \tag{3.4}$$

Note that from Theorem 1 of Fan and Wang (2007), using our notations, we can deduce immediately that the finite number of jumps in each log-price process are removed at a rate at least $(nL)^{-1/4}$ using the wavelet method, with nL being the total number of all-refresh data points. Individual asset may do even better since we use all data points available in practice for each asset before evaluating the all-refresh time points. This jump removal rate is in fact the key to the successful adaptation of wavelet jumps removal to our proposed nonlinear shrinkage estimator. More detailed assumptions:

(W1) The wavelets used in jump estimation are differentiable.

(W2) For the jump part of $X_t^{(j)}$ in $[0,1]$ for $j = 1,\ldots,p$, its jump locations $\eta_\ell^{(j)}$ and jump sizes $B_\ell^{(j)}$ satisfy

$$N_1^{(j)} < \infty,\ \eta_1^{(j)} < \ldots < \eta_\ell^{(j)} < \ldots,\ 0 < |B_\ell^{(j)}| < \infty \text{ almost surely.}$$

(W3) The number of stocks involved in a cojump is finite.

Assumptions (W1) and (W2) are technical assumptions adapted from Fan and Wang (2007). Assumption (W2) means that we are dealing with finite number of jumps for each log-price process, and the sizes of the jumps are bounded from 0 almost surely. If Assumption (W3) is not satisfied, then the rate of convergence of $\widehat{\mathbf{\Sigma}}(\tau_{j-1},\tau_j)$ in Theorem 2 using the jumps-removed data will be dependent on how many stocks is involved in a cojump in general. Our assumptions allow the jump process to be dependent on the drift, volatility and the microstructure noise process in general.

Theorem 4. *Let all the assumptions in Theorem 2 hold, as well as (W1) to (W3) for the jump-diffusion model (3.2). Using the jumps-removed all-refresh log-price data* $\mathbf{Y}^*(s) = \mathbf{Y}(s) - \widehat{\mathbf{J}}_{v_s},\ s = 1,\ldots,nL$ *in constructing the integrated covariance matrix estimator in (2.7), the same conclusions in Theorem 2 and Corollary 3 hold. Moreover, we have*

$$\Big\| \sum_{0\le t\le 1} (\Delta\mathbf{J}_t\Delta\mathbf{J}_t^{\mathrm{T}} - \Delta\widehat{\mathbf{J}}_t\Delta\widehat{\mathbf{J}}_t^{\mathrm{T}}) \Big\| = O_P(n^{-1/4}).$$

The following is the jumps-removal procedure:

1. Denote $Y_{i,k}^{(j)}$ the wavelet coefficients of $\{Y_t^{(j)}\}, k = 1,\ldots,2^i,\ i = 1,\ldots,\log_2(n),\ j = 1,\ldots,p.$
2. Let $D_n^{(j)} = d\sqrt{2\log n}$ be the universal threshold with d as the median of $|Y_{i_n,k}^{(j)}|$. If $|Y_{i_n,k}^{(j)}| > D_n^{(j)}$, the estimated jump location is $\widehat{\tau} = k2^{-i_n}$.

3. For a small neighbourhood δ_n of the estimated jump location, denote $\bar{Y}^{(j)}_{\widehat{\tau}_l+}$ and $\bar{Y}^{(j)}_{\widehat{\tau}_l-}$ as the average value over periods $[\widehat{\tau}_l, \widehat{\tau}_l + \delta_n]$ and $[\widehat{\tau}_l - \delta_n, \widehat{\tau}_l)$ respectively. We take δ_n as the square root of the total number of data points after data cleaning, following Fan and Wang (2007).

4. The estimated jump size is $\widehat{B}^{(j)}_l = \bar{Y}^{(j)}_{\widehat{\tau}_l+} - \bar{Y}^{(j)}_{\widehat{\tau}_l-}$, and the estimated jump variation is $\sum_{l=1}^{\widehat{q}} (\widehat{B}^{(j)}_l)^2$, where $\widehat{q}$ is the estimated number of jumps.

5. We remove the jump effect from the original observed data as $Y^{*(j)}_t = Y^{(j)}_t - \sum_{\widehat{\tau}_l \le t} \widehat{B}^{(j)}_l$.

3.2 Application to Portfolio Allocation

In this section we investigate the theoretical performance of our estimator when it is used to construct minimum-variance portfolios. Defining $\mathbf{1}_p$ as a column vector of p ones, we define the estimated optimal minimum-variance portfolio weights to be

$$\widehat{\mathbf{w}}_{\text{opt}} = \frac{\widehat{\mathbf{\Sigma}}(0,1)^{-1}\mathbf{1}_p}{\mathbf{1}_p^{\text{T}}\widehat{\mathbf{\Sigma}}(0,1)^{-1}\mathbf{1}_p},$$

where $\widehat{\mathbf{\Sigma}}(0,1)$ is our estimator of $\mathbf{\Sigma}(0,1)$. In Section 5, we empirically compare our estimator to other estimators using different measures, including the performance in minimizing portfolio risks.

Unlike DeMiguel et al. (2009) or Fan et al. (2012) which constrain the L_1 or L_2 norm of a portfolio vector $\mathbf{w}$ explicitly through a tuning parameter, our method enjoys a natural upper bound on the maximum exposure asymptotically in probability. The maximum exposure of a portfolio vector $\mathbf{w}$ is defined as $\|\mathbf{w}\|_{\max} = \max_i |w_i|$. The bound for our method is important since the theoretical minimum-variance portfolio is also subjected to the same bound. At the same time, the actual risk $R^{1/2}(\widehat{\mathbf{w}}_{\text{opt}}) = (\widehat{\mathbf{w}}_{\text{opt}}^{\text{T}}\mathbf{\Sigma}(0,1)\widehat{\mathbf{w}}_{\text{opt}})^{1/2}$ also has a natural upper bound, as presented below.

Theorem 5. *Let all the assumptions in Theorem 2 hold. Define the theoretical minimum variance portfolio weight to be*

$$\mathbf{w}_{\text{theo}} = \frac{\mathbf{\Sigma}(0,1)^{-1}\mathbf{1}_p}{\mathbf{1}_p^{\text{T}}\mathbf{\Sigma}(0,1)^{-1}\mathbf{1}_p}.$$

In the case of no pervasive factors with $p/n \to c > 0$, or the existence of pervasive factors with $p^{3/2}/n \to c > 0$, the maximum exposures of $\widehat{\mathbf{w}}_{\text{opt}}$ and $\mathbf{w}_{\text{theo}}$ satisfy, in probability,

$$p^{1/2}\|\widehat{\mathbf{w}}_{\text{opt}}\|_{\max},\ p^{1/2}\|\mathbf{w}_{\text{theo}}\|_{\max} \le \frac{\max_{1\le j\le L}\lambda_{\max}(\mathbf{\Sigma}(\tau_{j-1},\tau_j))}{\min_{1\le j\le L}\lambda_{\min}(\mathbf{\Sigma}(\tau_{j-1},\tau_j))}.$$

If there are no pervasive factors and $p/n \to c > 0$, the actual risks of $\widehat{\mathbf{w}}_{\text{opt}}$ and $\mathbf{w}_{\text{theo}}$ satisfy, in

probability,

$$p^{1/2}R^{1/2}(\widehat{\mathbf{w}}_{\text{opt}}) \leq \frac{\max_{1\leq j\leq L}\lambda_{\max}(\mathbf{\Sigma}(\tau_{j-1},\tau_j))}{\min_{1\leq j\leq L}\lambda_{\min}(\mathbf{\Sigma}(\tau_{j-1},\tau_j))}\cdot\lambda^{1/2}_{\max}(\mathbf{\Sigma}(0,1)),$$
$$p^{1/2}R^{1/2}(\mathbf{w}_{\text{theo}}) \leq \lambda^{1/2}_{\max}(\mathbf{\Sigma}(0,1)).$$

If there are pervasive factors and $p^{3/2}/n \to c > 0$, then $R(\widehat{\mathbf{w}}_{opt}) = O_P(\lambda_{\max}(\mathbf{\Sigma})) = O_P(p)$, where the bound for $R(\mathbf{w}_{theo})$ remains the same as above.

If Assumptions (W1) to (W3) hold also under the jump-diffusion model (3.2), then the same conclusions as above hold for the maximum exposure and actual risk bounds, as long as we are using the jumps-removed data as described in Section 3.1.

The proof of this theorem is in Section 7. The gross exposure constraint by Fan et al. (2012) or the L_2-norm constraint by DeMiguel et al. (2009) are useful in constraining the total exposure of a portfolio and obtaining special ones like the no-short-sale portfolio (by setting $\left\|\mathbf{w}\right\|_1 \leq 1$). In practice, as illustrated by our simulation experiments and real data analysis in Section 5, the maximum exposure can still be large while these explicit constraints are in place. Certainly, there are a lot of examples where concentrated portfolios can be rewarding. However, with respect to the minimum-variance portfolio, the theoretical one does satisfy an upper bound on the maximum exposure as presented in Theorem 5. Our method has the same upper bound in probability, which decays as p increases when there are no pervasive factors in the data. As illustrated in Section 5, the maximum exposure in $\widehat{\mathbf{w}}_{\text{opt}}$ is on average smaller than other state-of-the-art methods in various settings, especially when using a quarter of a trading day as a partition. At the same time, in the real data analysis, the risk for our method measured as the out-of-sample standard deviation of the return for a portfolio is smaller than all other methods in the two portfolio studies. The relatively small turnover of our portfolio as shown in Table 4 and 5 is also important when profitability is concerned. See Section 5 for more details.

When there are pervasive factors like the market factor in the data, we have $\left\|\widehat{\mathbf{w}}_{\text{opt}}\right\|_{\max} = O_P(p^{1/2}) = \left\|\mathbf{w}_{\text{theo}}\right\|_{\max}$ and $R(\left\|\widehat{\mathbf{w}}_{\text{opt}}\right\|) = O_P(p)$. It would seem that explicit constraints in the portfolio weights would be better than our method. However, the bounds are certainly not tight. Simulation results with pervasive factors in Table 3 show that our method still performs better than others with the smallest L_2 distance from the theoretical portfolio, and matches closely to its out-of-sample risk. It would need more sophisticated analysis to obtain tighter bounds when there are pervasive factors.

4 Practical Implementation

There are two parameters that can be tuned for potentially better performance, namely the partition $(\tau_{j-1}, \tau_j]$ of the period $[0, 1]$ (thus also determining L itself which represents the number of partitions), and the scale parameter K used in the TSCV in (2.4). For example, suppose we are given a period of 10 days of tick-by-tick data, if we set $(\tau_{j-1}, \tau_j]$ to be one day, then $L = 10$. Note that the length of each partition can be different. Similar to the function $g(m)$ in equation (4.7) of Lam (2016), we propose to minimize the following criterion for a good choice of $\boldsymbol{\tau} = \{\tau_j\}_{0 \le j \le L}$ and K:

$$g(\boldsymbol{\tau}, K) = \left\| \sum_{j=1}^{L} \left(\widehat{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j) - \widetilde{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j) \right) \right\|_F^2, \tag{4.1}$$

where $\widetilde{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j)$ and $\widehat{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j)$ are defined in (2.4) and (2.7) respectively. This function is inspired by Bickel and Levina (2008), where a similar function, with the population covariance matrix replaced by the sample covariance matrix, is used for the determination of the banding number in banding a large covariance matrix estimator. In our case, the above aligns with the optimization problem (2.6), but with $\boldsymbol{\Sigma}(\tau_{j-1}, \tau_j)$ replaced by the sample counterpart $\widetilde{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j)$. From our experience, as long as the intervals are not too different in length and that each interval has enough data points (at least the same order as p), the performance of the estimator is in fact more dependent on L, the number of partitions we choose. Hence we suggest to divide the time interval into equal length partitions, checking that each one has enough data points. We can then choose L by minimizing the criterion (4.1) above.

For the choice of K, since we are using $K \asymp n^{2/3}$ as in Zhang (2011), we can search $K = [bn^{2/3}]$ on a preset grid of constant b. In practice, we found from our simulation results and real data analysis that using $b = 1$ provide good results, and portfolio performance is not too different from using other values of b, hence in this paper we use $b = 1$.

5 Empirical Results

5.1 Simulation

In this section, we simulate high frequency trading transactions of 100 stocks for one year (250 trading days). The price processes and the asynchronous transaction times are simulated independently. The observed log-price is defined as $X_t^{o(i)} = X_t^{(i)} + \varepsilon_t^{(i)}$, where $X_t^{(i)}$ represents the latent log-price, and the microstructure noise has $\varepsilon_t^{(i)} \overset{iid}{\sim} N(0, 0.0005^2)$. We generate $p = 100$ latent log-prices by the following

Heston-like multivariate factor model with stochastic volatilities:

$$dX_t^{(i)} = \mu^{(i)}dt + \sqrt{1-(\rho^{(i)})^2}\sigma_t^{(i)}dB_t^{(i)} + \rho^{(i)}\sigma_t^{(i)}dW_t + C\nu^{(i)}dZ_t, \quad i = 1, \ldots, 100, \tag{5.1}$$

where $\{W_t\}$, $\{Z_t\}$ and the $\{B_t^{(i)}\}$'s are independent standard Brownian motions. The processes $\{W_t\}$ and $\{Z_t\}$ imitate factors in the market. The constant $C = 1_{\{\text{model 2}\}}$ is 0 for the first model we consider. We set $\rho^{(i)} = -0.7C$, so that it is 0 in the first model, and hence there are no factors. For the second model, $C = 1$, so that it contains two factors. The spot volatility $\sigma_t^{(i)} = \sqrt{\varrho_t^{(i)}}$ follows the Cox-Ingersoll-Ross (CIR) process

$$d\varrho_t^{(i)} = \kappa^{(i)}(\theta^{(i)} - \varrho_t^{(i)})dt + \xi^{(i)}dU_t^{(i)},$$

where the $\{U_t^{(i)}\}$'s are independent standard Brownian motions. Other parameters of $X_t^{(i)}$ are set at $(\mu^{(i)}, \kappa^{(i)}, \xi^{(i)}, \theta^{(i)}) = (0.03x_1^{(i)}, 1.1x_2^{(i)}, 0.5x_3^{(i)}, 0.25x_4^{(i)})$ and $\nu^{(i)} = \sqrt{\theta^{(i)}}$, where the $x_j^{(i)}$'s are independent uniform random variables on the interval $[0.7, 1.3]$. The initial value of each log-price $X_0^{(i)}$ is set randomly on the interval $[0.5, 1.5]$ and the starting spot volatility $\varrho_0^{(i)}$ on the interval $[0.5\theta^{(i)}, 1.5\theta^{(i)}]$.

For the transaction times, we generate 100 different Poisson processes with intensities $\lambda_1, \ldots, \lambda_{100}$ respectively. Since the normal trading time for one day is 23400 seconds, λ_i is set to be $0.01i \times 23400$, where $i = 1, \ldots, 100$.

5.2 Comparison of different estimators

5.2.1 Comparisons with TSCV and thresholded method

We compare our estimator to the TSCV, as well as the thresholded average realized volatility matrix (TARVM) which is essentially a thresholded TSCV introduced in Wang and Zou (2010). The reason we choose to compare to the TARVM on top of the TSCV is because when there are no factors, sparseness or approximate sparseness in $\mathbf{\Sigma}(0,1)$ can be natural as its eigenvalues are of constant order even with a diverging matrix dimension, giving potential advantages to thresholded estimators. Our estimator is a modified TSCV, and so comparing to another modified TSCV like the TARVM makes sense. Hereafter, we abbreviate our estimator as NERIVE when we are using one trading day as a partition length, and quarNERIVE when we are using a quarter of a trading day.

We use two measures for comparing the estimators. One is the Frobenius error, another is the average bias in eigenvalues, defined by

$$\text{Frobenius error} = \text{tr}(\widehat{\mathbf{\Sigma}}(0,1) - \mathbf{\Sigma}(0,1))^2, \quad \text{Average bias} = \text{tr}(\widehat{\mathbf{\Sigma}}(0,1) - \mathbf{\Sigma}(0,1))/p.$$

The integrated covariance matrix $\mathbf{\Sigma}(0,1)$ is evaluated using the simulated latent log-prices at the finest grid (1 per second). We divide the 250 trading days into disjoint 5-day intervals, and calculate the two error measures for different estimators over each 5-day interval. The means and standard deviations of these errors are reported in Table 1. It also includes the same exercise when 5-day becomes 1-day intervals. When we are using 5-day training windows, it is clear that NERIVE, especially quarNERIVE, performs

No factors ($C=0$)		NERIVE	quarNERIVE	TSCV	TARVM
5-day	Frobenius error	$12_{(1.3)}$	$7_{(0.9)}$	$156_{(22.2)}$	$64_{(5.3)}$
	Average bias	$30_{(1.6)}$	$23_{(1.4)}$	$33_{(2.3)}$	$36_{(1.6)}$
1-day	Frobenius error	-	$0.4_{(0.1)}$	$9.3_{(1.7)}$	$1.9_{(0.2)}$
	Average bias	-	$3_{(0.4)}$	$2_{(0.5)}$	$1_{(0.2)}$
With factors ($C=1$)		NERIVE	quarNERIVE	TSCV	TARVM
5-day	Frobenius error	$2007_{(1269)}$	$1161_{(539)}$	$3241_{(3370)}$	$3123_{(1527)}$
	Average bias	$59_{(14)}$	$45_{(8)}$	$67_{(25)}$	$72_{(14)}$
1-day	Frobenius error	-	$40_{(32)}$	$62_{(51)}$	$12_{(11)}$
	Average bias	-	$7_{(3.7)}$	$4_{(5.9)}$	$3_{(2.8)}$

Table 1: Mean and standard deviation of Frobenius error and average bias of eigenvalues over different 5-day or 1-day intervals for various methods. All values are multiplied by 10000.

better than TSCV and TAVRM in both measures. However, in using 1-day training windows, TAVRM is better in terms of average bias in the eigenvalues. When there are factors, TAVRM is also better in Frobenius norm error using 1-day training windows. It is clear that there are advantages in thresholding, especially when we consider a shorter window for the integrated covariance matrix, but our method is better in general when such window increases.

5.2.2 Comparisons with POET and related methods

POET (Principal Orthogonal complEment Thresholding), originally proposed as a general low frequency data method in Fan et al. (2013), essentially assumes that the true covariance matrix can be decomposed into a low rank matrix (induced from factors in the data) plus a sparse residual one. Aït-Sahalia and Xiu (2017) proposes such a decomposition on the realized covariance matrix of subsampled return data (15 or 30 minutes interval) to reduce the effects of microstructure noise contamination, while the residual covariance is assumed to be block diagonal with known blocks (e.g., blocking by industry). Dao et al. (2017) proposes the POET method on realized covariance matrix calculated on pre-averaged return data (PRVM), with thresholding developed for the residual matrix. We find that such thresholding usually works better than blocking using industry, and so we compare our method to such a POET method applied on TSCV (TS-POET), since NERIVE or quarNERIVE are based on nonlinear shrinkage of the TSCV.

We also explore if our nonlinear shrinkage can be applied to the PRVM rather than TSCV. To be precise, we replace $\widetilde{\mathbf{\Sigma}}(\tau_{\ell-1},\tau_\ell)$ in (2.4) by the corresponding PRVM, and follow Section 2.2 to construct $\widehat{\mathbf{\Sigma}}(0,1)$ in (2.8). We abbreviate nonlinear shrinkage based on PRVM as PR-NERIVE or PR-quarNERIVE, and

compare them with the POET in Dao et al. (2017) (PR-POET). Since Fan and Kim (2017) has developed a robust version of PRVM with POET (RPR-POET), we compare this to PR-NERIVE and PR-quadNERIVE as well. Throughout the rest of the paper, all POET methods use 5 factors which is enough to achieve consistently good results.

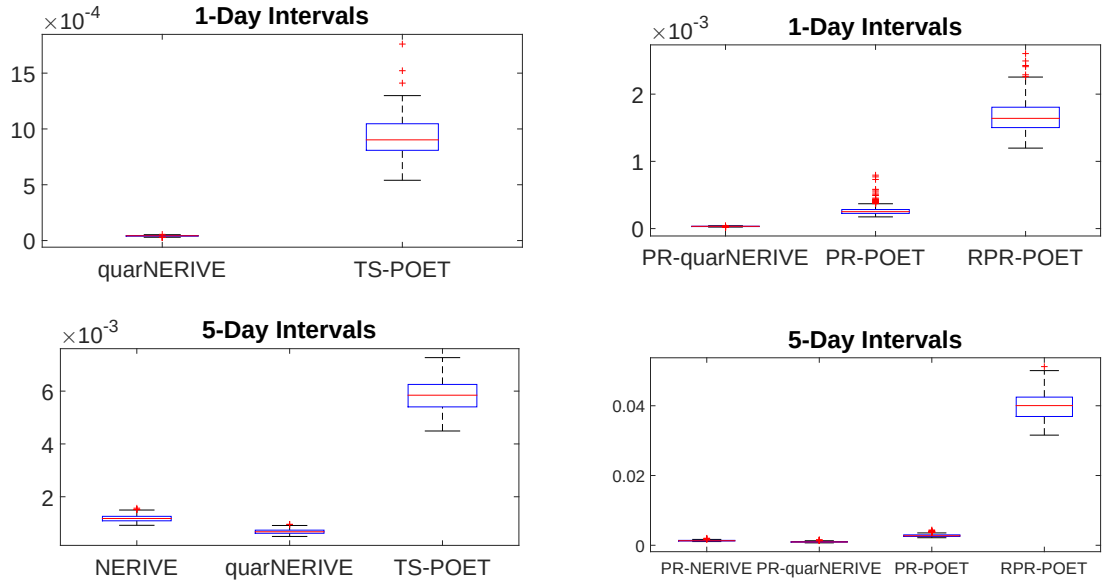

Figure 1: Boxplot of Frobenius errors when there are no factors in model (5.1) ($C = 0$).

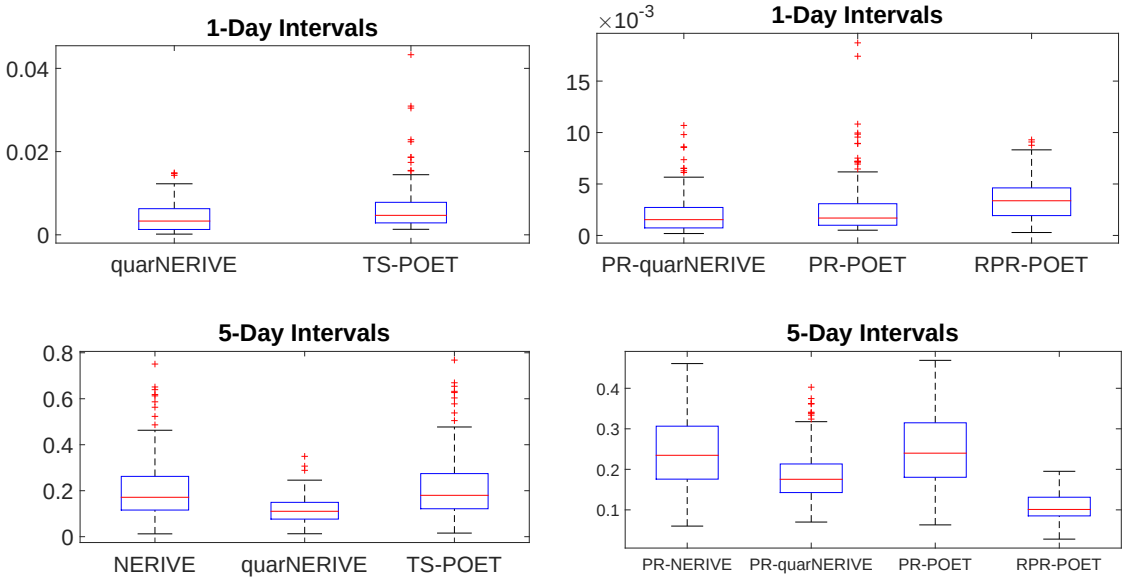

Figure 2: Boxplot of Frobenius errors when there are factors in model (5.1) ($C = 1$).

Figure 1 shows the Frobenius errors when there are no factors in model (5.1). It is clear that quarNERIVE is better than NERIVE and TS-POET. When we use pre-averaged data for nonlinear shrinkage, PR-quarNERIVE is also better than PR-NERIVE, PR-POET and RPR-POET. When there are factors in model (5.1), quarNERIVE is still better than NERIVE and TS-POET, but PR-quarNEIVE is not as good as RPR-POET when we consider a longer time horizon for the integrated covariance matrix (5-day). Clearly PR-NERIVE has a lot of potential, and we hope to develop its theoretical performance in another project (see Remark 1 as well). We have also considered the spectral error, but the patterns are very similar to Figure 1 and 2, and hence they are omitted.

5.3 Comparison of portfolio allocation performance

To compare the performance of different methods, we focus on the minimum variance portfolio

$$\mathbf{w}_{\text{opt}} = \frac{\boldsymbol{\Sigma}(0,1)^{-1}\mathbf{1}_p}{\mathbf{1}_p^{\mathrm{T}}\boldsymbol{\Sigma}(0,1)^{-1}\mathbf{1}_p}, \quad \text{which solves} \min_{\mathbf{w}:\mathbf{w}^{\mathrm{T}}\mathbf{1}_p=1} \mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}(0,1)\mathbf{w}.$$

We first set the benchmark for comparisons. Following Fan et al. (2012), we create a theoretical portfolio $\mathbf{w}_{\text{theo}}$, which is a minimum variance portfolio with $\boldsymbol{\Sigma}(0,1)$ evaluated similarly as in Section 5.2. For all other methods, we use the all-refresh time points evaluated from the data (we do not hold positions overnight for all methods to avoid overnight price jumps, since they are not what our study is about).

Other portfolios are constructed and compared to the theoretical minimum variance portfolio (THEO) above. The first one is the equal weight portfolio (EQUAL). The second one is the minimum variance portfolio with $\boldsymbol{\Sigma}(0,1)$ substituted by the two scale covariance matrix (TSCV). We abbreviate it as TARVM when $\boldsymbol{\Sigma}(0,1)$ is replaced by the TARVM as in Section 5.2. When $\boldsymbol{\Sigma}(0,1)$ is substituted with our estimator, we abbreviate it as NERIVE with one trading day as a partition length, and quarNERIVE when a partition length is a quarter of a trading day. We also compare with the gross exposure constraint (GEC) method (Fan et al., 2012), and the L_2 norm constraint (NORM) (DeMiguel et al., 2009) based on TSCV. The GEC and NORM methods solve respectively

$$\text{GEC}: \min_{\mathbf{w}:\mathbf{w}^{\mathrm{T}}\mathbf{1}_p=1, \left\|\mathbf{w}\right\|_1 \le c} \mathbf{w}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}(0,1)\mathbf{w},$$

$$\text{NORM}: \min_{\mathbf{w}:\mathbf{w}^{\mathrm{T}}\mathbf{1}_p=1, \left\|\mathbf{w}\right\|_2^2 \le \delta} \mathbf{w}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}(0,1)\mathbf{w}.$$

We constructed 3 GEC portfolios with tuning parameters $c = 1, 2, 3$, as well as 3 NORM portfolios with tuning parameters $\delta = 0.1, 0.5, 1$ for comparisons. We do not use the pairwise refresh method for GEC

to save significant computational time in both the simulations and the real data analysis, as well as that the features of our method can be compared more directly to those of GEC. Finally, we also compare to TS-POET, PR-NERIVE, PR-quarNERIVE, PR-POET and RPR-POET as in Section 5.2.2 when the corresponding estimator substitutes $\mathbf{\Sigma}(0,1)$ in $\mathbf{w}_{\text{opt}}$.

The portfolio exercise is carried out as follows for all methods. We invest 1 unit of capital to the different portfolios above at a certain start date (e.g., day 6 if we are using a 5-day training window), and rebalance the portfolio weights daily, moving the training window one day ahead. There are two investment strategies for comparisons under each model 1 or 2. The first one rebalances the portfolio daily with a 5-day training window. The second one rebalances the portfolio daily with a 1-day training window.

The quantities to be compared for different portfolios are as follows. For daily rebalancing with a k-day training window ($k = 1$ or 5), we calculate the annualized portfolio return and annualized out-of-sample standard deviation, given respectively by

$$\widehat{\mu} = 250 \times \frac{1}{250-k} \sum_{i=k+1}^{250} \mathbf{w}^{\mathrm{T}} \mathbf{r}_i, \quad \widehat{\sigma} = \Big(250 \times \frac{1}{250-k} \sum_{i=k+1}^{250} (\mathbf{w}^{\mathrm{T}} \mathbf{r}_i - \frac{\widehat{\mu}}{250})^2\Big)^{1/2}.$$

The out-of-sample standard deviation is a good indicator of how much risk is associated with a portfolio (DeMiguel et al., 2009), and is our main quantity for performance comparisons, whereas portfolio return is of secondary importance. We also calculate the Sharpe ratio $\widehat{\mu}/\widehat{\sigma}$. The average maximum exposure and the maximum of the maximum exposure over the whole investment horizon are two important measures for comparisons too. Since this is a simulation experiment, we can calculate the actual risk of a portfolio $\mathbf{w}$, $R^{1/2}(\mathbf{w}) = (\mathbf{w}^{\mathrm{T}} \mathbf{\Sigma} \mathbf{w})^{1/2}$, over a trading day. We compare the averaged actual risks of different methods over the whole investment horizon. Finally we compare the error norm compared to $\mathbf{w}_{\text{theo}}$, defined as $\text{Norm} = \big\|\mathbf{w} - \mathbf{w}_{\text{theo}}\big\|$, and also the portfolio turnover for different methods.

Table 2 shows the results for model (5.1) with no factors. Excluding all methods based on pre-averaged return data, the out-of-sample standard deviations of NERIVE and quarNERIVE are among the smallest for both 5-day and 1-day training windows, and closely match that of the theoretical minimum portfolio. TS-POET is the best when we are using 1-day training window. Pre-averaging tends to improve on nonlinear shrinkage and POET also, with PR-POET the best when we are using 1-day training window. The equal weight portfolio performs well also but is not as good as our methods when we use 5-day training windows. Our methods also have (together with TS-POET and PR-POET) among the closest L_2 distance from the theoretical minimum portfolio weight, and apart from GEC1, PR-quarNERIVE has the smallest portfolio turnover. Both TSCV and TARVM are having much larger actual risks than other methods, and a lot of times with impractical maximum exposures.

Table 3 shows the results for model (5.1) with factors. In general, risks are higher with factors, even for the theoretical portfolio. Our methods (quarNERIVE or PR-quarNERIVE) have risks close to the theoretical ones, with portfolio weights the closest to the theoretical portfolio weights among all methods. Equal weight portfolio now performs at a similar level to other methods (apart from TSCV and TARVM) in terms of risk minimization, but our methods are around 50% better in minimizing the our-of-sample SD or the actual risk. TSCV and TARVM are still the worst in terms of risks, maximum exposures and portfolio turnover. Overall, NERIVE or quarNERIVE (and their pre-averaging versions) do well in risk minimization compared to all other methods including the equal weight portfolio, with reasonable and often small maximum exposures and portfolio turnover.

Methods	Out-of Sample SD (%)	Actual Risk (%)	Norm	Average Max Abs Wgt(%)	Max Max Abs Wgt(%)	Portfolio Turnover	Portfolio Return (%)	Sharpe Ratio
daily rebalancing portfolio with 5-day training window								
THEO	1.6	1.7	–	$10_{(6.2)}$	44	$0.06_{(0.02)}$	5.2	3.2
NERIVE	1.9	1.9	0.08	$6_{(2.7)}$	18	$0.14_{(0.02)}$	7.4	3.9
quarNERIVE	1.8	1.9	0.07	$7_{(3.0)}$	19	$0.12_{(0.02)}$	6.0	3.3
EQUAL	2.0	2.2	0.13	$1_{(-)}$	1	–	5.5	2.8
TSCV	149.6	149.2	1.34	$64_{(352.7)}$	5066	$4.21_{(33.43)}$	297.5	2.0
GEC1	2.1	2.3	0.13	$2_{(1.0)}$	7	$0.06_{(0.04)}$	6.1	2.9
GEC2	2.5	2.6	0.13	$8_{(4.0)}$	32	$0.35_{(0.06)}$	3.9	1.6
GEC3	2.9	2.9	0.15	$7_{(2.1)}$	14	$0.42_{(0.08)}$	4.5	1.5
NORM0.1	3.6	3.5	0.18	$8_{(3.6)}$	24	$0.62_{(0.24)}$	9.8	2.7
NORM0.5	5.8	5.3	0.28	$16_{(8.9)}$	54	$1.20_{(0.51)}$	2.0	0.3
NORM1	7.1	6.6	0.36	$20_{(13.8)}$	80	$1.56_{(0.92)}$	7.9	1.1
TARVM	7.9	14.5	0.44	$34_{(119.6)}$	1276	$1.47_{(4.27)}$	−21.6	−2.7
TS-POET	2.0	2.0	0.08	$9_{(4.1)}$	27	$0.15_{(0.03)}$	9.7	4.9
PR-NERIVE	1.8	1.9	0.08	$6_{(2.6)}$	14	$0.11_{(0.02)}$	6.6	3.6
PR-quarNERIVE	1.8	1.9	0.08	$6_{(2.5)}$	13	$0.10_{(0.02)}$	6.0	3.3
PR-POET	1.8	1.9	0.06	$9_{(4.3)}$	27	$0.10_{(0.02)}$	6.0	3.4
RPR-POET	5.6	6.0	0.32	$26_{(7.3)}$	52	$0.36_{(0.15)}$	1.0	0.2
daily rebalancing portfolio with 1-day training window								
THEO	1.6	1.7	–	$10_{(6.2)}$	44	$0.06_{(0.02)}$	5.7	3.6
quarNERIVE	2.1	2.2	0.13	$2_{(0.3)}$	4	$0.2_{(0.03)}$	5.4	2.6
EQUAL	2.0	2.2	0.13	$1_{(-)}$	1	–	6.1	3.1
TSCV	373.4	806.1	5.43	$172_{(1187.5)}$	17696	$15.95_{(117.09)}$	1204.9	3.2
GEC1	2.0	2.2	0.13	$1_{(0.1)}$	2	$0.04_{(0.01)}$	6.0	3.0
GEC2	6.9	7.1	0.29	$11_{(6.6)}$	36	$1.02_{(0.32)}$	−5.0	−0.7
GEC3	16.5	16.1	0.63	$28_{(11.1)}$	58	$2.38_{(1.10)}$	−45.6	−2.8
NORM0.1	6.6	6.5	0.28	$7_{(2.8)}$	13	$1.28_{(0.31)}$	2.8	0.4
NORM0.5	14.8	14.7	0.60	$19_{(6.0)}$	32	$3.38_{(1.14)}$	−0.5	0.0
NORM1	20.7	19.3	0.80	$25_{(8.1)}$	51	$-12.41_{(268.64)}$	2.5	0.1
TARVM	342.8	511.0	4.64	$139_{(652.4)}$	7949	$14.63_{(144.97)}$	902.8	2.6
TS-POET	1.8	2.0	0.08	$9_{(4.4)}$	28	$0.29_{(0.03)}$	3.3	1.8
PR-quarNERIVE	2.0	2.2	0.12	$2_{(0.3)}$	3	$0.17_{(0.02)}$	5.3	2.6
PR-POET	1.7	1.9	0.06	$9_{(5.4)}$	35	$0.21_{(0.03)}$	5.7	3.3
RPR-POET	5.8	6.2	0.31	$25_{(9.1)}$	66	$0.98_{(0.32)}$	-1.4	-0.2

Table 2: Simulation results for model 1 with no factors ($C = 0$ in (5.1)): Annualized out-of-sample standard deviation, actual risk, norm of weights difference, averaged maximum absolute weight (standard deviation in bracket), maximum of maximum absolute weight, portfolio return and Sharpe ratio for various methods, including GEC ($c = 1, 2, 3$) and NORM ($\delta = 0.1, 0.5, 1$).

Methods	Out-of Sample SD (%)	Actual Risk (%)	Norm	Average Max Abs Wgt(%)	Max Max Abs Wgt(%)	Portfolio Turnover	Portfolio Return (%)	Sharpe Ratio
daily rebalancing portfolio with 5-day training window								
THEO	13	13	–	$41_{(21)}$	143	$0.3_{(0.1)}$	−1.5	−0.1
NERIVE	14	15	0.55	$22_{(5)}$	42	$0.9_{(0.2)}$	0.8	0.1
quarNERIVE	14	14	0.53	$23_{(6)}$	48	$0.8_{(0.2)}$	−10.7	−0.8
EQUAL	27	27	0.97	$1_{(-)}$	1	–	28.8	1.1
TSCV	34235	19343	107.02	$3937_{(51415)}$	804861	$468.0_{(6900.5)}$	−75911.4	−2.2
GEC1	25	25	1.00	$45_{(15)}$	94	$0.5_{(0.2)}$	20.2	0.8
GEC2	24	24	1.00	$47_{(16)}$	98	$0.7_{(0.3)}$	20.3	0.9
GEC3	23	24	1.01	$47_{(15)}$	106	$0.9_{(0.5)}$	14.8	0.6
NORM0.1	21	22	0.87	$8_{(1)}$	14	$0.6_{(0.2)}$	5.5	0.3
NORM0.5	17	18	0.78	$18_{(2)}$	26	$1.2_{(0.4)}$	−44.7	−2.6
NORM1	18	18	0.86	$28_{(3)}$	44	$1.7_{(0.6)}$	−49.2	−2.8
TARVM	8777	13790	91.58	$3077_{(33547)}$	524587	$61.7_{(447.5)}$	18387.1	2.1
TS-POET	17	17	0.77	$31_{(7)}$	59	$1.8_{(0.3)}$	-5.9	-0.4
PR-NERIVE	14	14	0.53	$21_{(5)}$	42	$0.9_{(0.1)}$	-4.7	-0.3
PR-quarNERIVE	14	14	0.52	$22_{(5)}$	39	$0.8_{(0.1)}$	-11.5	-0.8
PR-POET	15	15	0.57	$32_{(8)}$	60	$1.2_{(0.2)}$	-11.2	-0.7
RPR-POET	19	20	0.96	$41_{(10)}$	81	$2.4_{(0.7)}$	-16.9	-0.9
daily rebalancing portfolio with 1-day training window								
THEO	13	13	–	$41_{(21)}$	143	$0.3_{(0.1)}$	0.4	0.0
quarNERIVE	16	17	0.75	$17_{(2)}$	26	$2.0_{(0.6)}$	−37.5	−2.3
EQUAL	27	27	0.97	$1_{(-)}$	1	–	20.8	0.8
TSCV	1605	2061	30.52	$1054_{(5474)}$	77945	$39.7_{(439.9)}$	112.1	0.1
GEC1	26	27	1.05	$44_{(17)}$	95	$0.8_{(0.2)}$	−6.0	−0.2
GEC2	26	26	1.05	$46_{(16)}$	103	$1.1_{(0.3)}$	−12.8	−0.5
GEC3	25	25	1.06	$48_{(16)}$	106	$1.6_{(0.6)}$	−27.1	−1.1
NORM0.1	22	23	0.89	$8_{(1)}$	14	$1.4_{(0.2)}$	4.2	0.2
NORM0.5	21	22	0.95	$19_{(3)}$	33	$3.1_{(1.0)}$	3.4	0.2
NORM1	26	26	1.23	$29_{(5)}$	52	$5.7_{(5.3)}$	22.2	0.8
TARVM	2707	2714	23.87	$667_{(4307)}$	67297	$1.3_{(536.7)}$	5567.6	2.1
TS-POET	20	20	0.84	$22_{(6)}$	51	$2.5_{(0.3)}$	-1.1	-0.1
PR-quarNERIVE	16	17	0.72	$16_{(2)}$	25	$2.0_{(0.2)}$	-39.9	-2.5
PR-POET	17	18	0.76	$27_{(8)}$	56	$2.7_{(0.3)}$	11.7	0.7
RPR-POET	23	24	1.21	$51_{(22)}$	159	$5.1_{(2.2)}$	64.7	2.8

Table 3: Simulation results for model 2 with factors ($C = 1$ in (5.1)): Annualized out-of-sample standard deviation, actual risk, norm of weights difference, averaged maximum absolute weight (standard deviation in bracket), maximum of maximum absolute weight, portfolio return and Sharpe ratio for various methods, including GEC ($c = 1, 2, 3$) and NORM ($\delta = 0.1, 0.5, 1$).

5.4 Portfolio allocation study - real data analysis

In this study, we choose the stocks based on two lists, the "Fifty Most Active Stocks on NYSE, Round Lots (mils. of shares), 2013" and "Fifty Most Active Stocks by Dollar Volume on NYSE ($ in mils.), 2013", from the New York Stock Exchange Data official website `http://www.nyxdata.com/`. There are 26 stocks appearing in both of the lists above, and 74 stocks in either of them. We downloaded all the trading transactions of these 74 stocks in Year 2013 from the Wharton Research Data Services (WRDS, `https://wrds-web.wharton.upenn.edu/`). We omit the stock Sprint Corporation due to missing price data. We first clean all the data by the R-package "highfrequency", which follows the high frequency data cleaning steps presented in Barndorff-Nielsen et al. (2009). We conduct our portfolio allocation study on two portfolios, one with the $p = 26$ stocks appearing in both lists, and the other with $p = 73$ stocks appearing in either of the lists.

p = 26 Methods	Out-of-Sample SD (%)	Aver Max Abs Weight(%)	Max Max Abs Weight(%)	Portfolio Turnover	Portfolio Return(%)	Sharpe Ratio
daily rebalancing portfolio with 5-day training window						
NERIVE	4.5	$21_{(6)}$	41	$0.26_{(0.1)}$	18.6	4.2
quarNERIVE	4.4	$20_{(5)}$	36	$0.22_{(0.1)}$	21.0	4.8
EQUAL	5.2	$4_{(-)}$	4	$-$	24.3	4.6
TSCV	6.1	$42_{(13)}$	84	$1.16_{(0.5)}$	16.9	2.8
GEC1	5.0	$30_{(11)}$	69	$0.33_{(0.1)}$	28.0	5.6
GEC2	4.9	$34_{(11)}$	78	$0.58_{(0.2)}$	20.1	4.1
GEC3	5.4	$39_{(12)}$	83	$0.88_{(0.3)}$	18.0	3.3
NORM0.1	4.6	$13_{(2)}$	19	$0.25_{(0.1)}$	18.6	4.0
NORM0.5	5.4	$33_{(6)}$	52	$0.84_{(0.3)}$	14.3	2.7
NORM1	5.9	$41_{(11)}$	74	$1.09_{(0.4)}$	14.7	2.5
TARVM	13.3	$56_{(75)}$	1097	$1.74_{(3.0)}$	-9.1	-0.7
TS-POET	5.0	$30_{(8)}$	59	$0.60_{(0.2)}$	18.9	3.7
PR-NERIVE	4.4	$19_{(5)}$	39	$0.23_{(0.1)}$	18.1	4.1
PR-quarNERIVE	4.3	$19_{(5)}$	36	$0.22_{(0.1)}$	19.4	4.5
PR-POET	4.4	$25_{(7)}$	48	$0.34_{(0.1)}$	18.2	4.1
RPR-POET	8.0	$50_{(15)}$	110	$1.12_{(0.4)}$	21.0	2.6
daily rebalancing portfolio with 1-day training window						
quarNERIVE	4.5	$19_{(7)}$	42	$0.58_{(0.2)}$	17.8	3.9
EQUAL	5.2	$4_{(-)}$	4	$-$	24.1	4.6
TSCV	7.0	$60_{(80)}$	973	$3.15_{(7.3)}$	19.6	2.8
GEC1	4.9	$28_{(13)}$	67	$0.54_{(0.2)}$	28.2	5.7
GEC2	5.0	$33_{(12)}$	74	$0.97_{(0.2)}$	26.7	5.3
GEC3	5.5	$36_{(13)}$	87	$1.32_{(0.3)}$	24.5	4.4
NORM0.1	4.6	$12_{(2)}$	19	$0.52_{(0.1)}$	20.7	4.5
NORM0.5	5.3	$29_{(6)}$	47	$1.33_{(0.2)}$	18.3	3.5
NORM1	5.8	$41_{(10)}$	70	$1.86_{(0.4)}$	17.2	3.0
TARVM	13.4	$40_{(93)}$	878	$3.49_{(18.5)}$	3.4	0.3
TS-POET	4.8	$28_{(11)}$	85	$1.24_{(0.4)}$	22.7	4.7
PR-quarNERIVE	4.5	$18_{(6)}$	38	$0.55_{(0.2)}$	16.5	3.7
PR-POET	4.8	$26_{(9)}$	59	$1.00_{(0.3)}$	22.1	4.6
RPR-POET	9.2	$67_{(27)}$	193	$2.50_{(0.6)}$	-2.2	-0.2

Table 4: Empirical results for the 26 most actively traded stocks in NYSE: annualized out-of-sample standard deviation, averaged maximum absolute weight, maximum of maximum absolute weight, annualized portfolio return and Sharpe ratio for various methods, including GEC with $c = 1, 2, 3$, and NORM with $\delta = 0.1, 0.5, 1$.

We carry out the same portfolio allocation exercises as in our simulations for both the 26-stock and 73-stock portfolios. First we do not remove jumps from the cleaned data. The results are displayed in Tables 4 and 5. Both NERIVE and quarNERIVE achieve the lowest out-of-sample SD in the two scenarios presented for both portfolios, which are all under 4.5%. PR-POET has similarly good performance too, although with higher maximum exposure compared to NERIVE, quarNERIVE and their pre-averaging versions. The maximum exposure of our methods are reasonably low even compared to the no-short-sale or L_2-constrained portfolios, with among the lowest portfolio turnovers under all scenarios for both $p = 26$ and $p = 73$ portfolios.

We also considered jumps removed data. The results are presented in Table 6 and Table 7. In general, the out-of-sample SD do not change much for all methods, except for TSCV and TARVM which can see huge increase or decrease in the risk. It is not surprising though as both methods can invest heavily in all stocks, rendering them more sensitive to jumps removal. In fact the number of jumps estimated for each

p = 73 Methods	Out-of-Sample SD (%)	Aver Max Abs Weight(%)	Max Max Abs Weight(%)	Portfolio Turnover	Portfolio Return(%)	Sharpe Ratio
daily rebalancing portfolio with 5-day training window						
NERIVE	3.8	$12_{(3)}$	22	$0.44_{(0.1)}$	15.3	4.0
quarNERIVE	3.9	$12_{(3)}$	21	$0.40_{(0.1)}$	16.1	4.1
EQUAL	5.4	$1_{(-)}$	1	$-$	22.3	4.1
TSCV	470.8	$629_{(4042)}$	58950	$104.71_{(1554.4)}$	1367.1	2.9
GEC1	5.0	$21_{(11)}$	57	$0.34_{(0.2)}$	21.4	4.3
GEC2	4.7	$25_{(10)}$	64	$0.57_{(0.2)}$	14.2	3.0
GEC3	4.7	$25_{(9)}$	59	$0.87_{(0.2)}$	9.5	2.0
NORM0.1	4.5	$9_{(1)}$	15	$0.46_{(0.1)}$	14.7	3.3
NORM0.5	4.9	$19_{(5)}$	33	$1.29_{(0.3)}$	8.8	1.8
NORM1	5.8	$26_{(7)}$	48	$2.07_{(0.6)}$	7.9	1.4
TARVM	4.6	$5_{(1)}$	12	$0.09_{(0.0)}$	21.5	4.7
TS-POET	4.2	$19_{(5)}$	37	$0.87_{(0.3)}$	14.6	3.4
PR-NERIVE	3.9	$11_{(3)}$	21	$0.43_{(0.1)}$	15.7	4.0
PR-quarNERIVE	3.9	$11_{(3)}$	22	$0.39_{(0.1)}$	15.3	3.9
PR-POET	3.8	$16_{(5)}$	32	$0.56_{(0.2)}$	17.0	4.4
RPR-POET	5.6	$15_{(4)}$	33	$0.61_{(0.2)}$	22.6	4.0
daily rebalancing portfolio with 1-day training window						
quarNERIVE	4.2	$8_{(3)}$	20	$0.79_{(0.2)}$	18.6	4.4
EQUAL	5.4	$1_{(-)}$	1	$-$	22.4	4.2
TSCV	120.3	$381_{(671)}$	3897	$25.80_{(214.4)}$	-33.2	-0.3
GEC1	5.1	$7_{(13)}$	69	$0.17_{(0.2)}$	19.2	3.7
GEC2	5.5	$19_{(12)}$	61	$0.83_{(0.2)}$	25.3	4.6
GEC3	9.7	$30_{(27)}$	158	$1.58_{(3.9)}$	29.8	3.1
NORM0.1	4.6	$8_{(3)}$	19	$0.76_{(0.2)}$	21.4	4.6
NORM0.5	8.0	$18_{(10)}$	50	$1.95_{(0.9)}$	29.7	3.7
NORM1	11.0	$27_{(17)}$	69	$3.25_{(3.5)}$	21.4	1.9
TARVM	103.9	$221_{(455)}$	4726	$44.63_{(324.1)}$	136.7	1.3
TS-POET	5.2	$23_{(11)}$	83	$1.78_{(0.5)}$	15.8	3.0
PR-quarNERIVE	4.3	$8_{(2)}$	18	$0.79_{(0.2)}$	16.9	3.9
PR-POET	4.2	$13_{(5)}$	40	$0.94_{(0.3)}$	20.1	4.8
RPR-POET	6.4	$17_{(5)}$	43	$1.10_{(0.2)}$	23.7	3.7

Table 5: Empirical results for the 73 most actively traded stocks in NYSE: annualized out-of-sample standard deviation, averaged maximum absolute weight, maximum of maximum absolute weight, annualized portfolio return and Sharpe ratio for various methods, including GEC with $c = 1, 2, 3$, and NORM with $\delta = 0.1, 0.5, 1$.

date is typically around 4 or 5, which is a very small number compared to the number of all-refresh data points.

6 Conclusion

We generalize nonlinear shrinkage of eigenvalues in a large sample covariance matrix for independent and identically distributed random vectors (Lam, 2016) to that of a large two-scale covariance matrix estimator (TSCV) for high frequency returns, which are not independent and identically distributed in general. To do this, we split the data into partitions and regularize the eigenvalues of the TSCV within a partition by the data from other partitions. Regularization is indeed achieved both theoretically and empirically, as demonstrated by the good performance in our simulations and portfolio allocation exercises.

Since TSCV has a slower rate of convergence than the multi-scale realized volatility matrix (Tao et al.,

p = 26 Methods	Out-of-Sample SD (%)	Aver Max Abs Weight(%)	Max Max Abs Weight(%)	Portfolio Turnover	Portfolio Return(%)	Sharpe Ratio
daily rebalancing portfolio with 5-day training window						
NERIVE	4.5	$20_{(6)}$	44	$0.26_{(0.1)}$	17.1	3.8
quarNERIVE	4.4	$19_{(5)}$	34	$0.23_{(0.1)}$	19.2	4.4
EQUAL	5.2	$4_{(-)}$	4	$-$	24.3	4.6
TSCV	5.9	$41_{(13)}$	92	$1.09_{(0.4)}$	16.5	2.8
GEC1	5.0	$30_{(11)}$	66	$0.33_{(0.1)}$	29.9	6.0
GEC2	5.0	$35_{(10)}$	80	$0.68_{(0.2)}$	19.0	3.8
GEC3	5.5	$39_{(11)}$	88	$0.95_{(0.3)}$	17.6	3.2
NORM0.1	4.6	$13_{(2)}$	18	$0.26_{(0.1)}$	17.6	3.8
NORM0.5	5.3	$33_{(6)}$	52	$0.83_{(0.3)}$	15.8	3.0
NORM1	5.7	$40_{(11)}$	76	$1.04_{(0.4)}$	15.8	2.7
TARVM	6.7	$51_{(37)}$	550	$1.53_{(1.6)}$	17.8	2.7
TS-POET	5.1	$30_{(8)}$	58	$0.60_{(0.2)}$	16.9	3.3
PR-NERIVE	4.4	$19_{(5)}$	34	$0.23_{(0.1)}$	20.1	4.6
PR-quarNERIVE	4.4	$20_{(5)}$	35	$0.22_{(0.1)}$	20.9	4.8
PR-POET	4.3	$25_{(6)}$	50	$0.33_{(0.1)}$	17.0	3.9
RPR-POET	7.9	$50_{(14)}$	110	$1.14_{(0.4)}$	17.5	2.2
daily rebalancing portfolio with 1-day training window						
quarNERIVE	4.5	$18_{(7)}$	41	$0.58_{(0.2)}$	17.0	3.7
EQUAL	5.2	$4_{(-)}$	4	$-$	24.1	4.6
TSCV	6.5	$55_{(49)}$	708	$2.54_{(1.8)}$	23.4	3.6
GEC1	4.9	$28_{(12)}$	62	$0.55_{(0.2)}$	29.2	6.0
GEC2	5.0	$33_{(13)}$	84	$0.95_{(0.2)}$	28.5	5.8
GEC3	5.4	$36_{(13)}$	85	$1.32_{(0.3)}$	26.9	5.0
NORM0.1	4.6	$12_{(2)}$	20	$0.52_{(0.1)}$	20.9	4.5
NORM0.5	5.2	$30_{(6)}$	62	$1.34_{(0.2)}$	16.6	3.2
NORM1	5.7	$41_{(10)}$	73	$1.86_{(0.4)}$	17.8	3.1
TARVM	14.2	$42_{(90)}$	724	$0.38_{(20.6)}$	3.3	0.2
TS-POET	4.9	$28_{(11)}$	77	$1.25_{(0.4)}$	21.2	4.3
PR-quarNERIVE	4.5	$18_{(6)}$	40	$0.56_{(0.2)}$	17.3	3.8
PR-POET	4.7	$26_{(9)}$	63	$1.01_{(0.3)}$	23.7	5.0
RPR-POET	9.2	$65_{(27)}$	214	$2.51_{(0.7)}$	-9.2	-1.0

Table 6: Empirical results (jumps removed) for the 26 most actively traded stocks in NYSE: annualized out-of-sample standard deviation, averaged maximum absolute weight, maximum of maximum absolute weight, annualized portfolio return and Sharpe ratio for various methods, including GEC with $c = 1, 2, 3$, and NORM with $\delta = 0.1, 0.5, 1$.

2013), the kernel realized volatility matrix (Barndorff-Nielsen et al., 2011b) or the pre-averaging realized volatility positive semi-definite matrix (Christensen et al., 2010), there are potential improvements if our method is applied to these estimators. Indeed, simulation and empirical results in Section 5 do suggest that pre-averaging can improve nonlinear shrinkage performance further. Comparisons with the thresholded version of these estimators (Kim et al., 2016) will also be revealing, and we leave these works in a future project.

7 Proof of Theorems

Before presenting the proofs, we present the last set of assumptions which are required for Theorem 2 to hold. We first need to decompose $\mathbf{X}_{v_s} - \mathbf{X}(s)$. Consider the previous-tick time $t_s^i \in (v_{s-1}, v_s]$ for the ith

p = 73 Methods	Out-of-Sample SD (%)	Aver Max Abs Weight(%)	Max Max Abs Weight(%)	Portfolio Turnover	Portfolio Return(%)	Sharpe Ratio
daily rebalancing portfolio with 5-day training window						
NERIVE	3.9	$12_{(3)}$	22	$0.44_{(0.1)}$	14.9	3.9
quarNERIVE	3.9	$12_{(3)}$	21	$0.39_{(0.1)}$	15.3	3.9
EQUAL	5.4	$1_{(-)}$	1	–	22.3	4.1
TSCV	126.6	$236_{(503)}$	4693	$25.31_{(96.7)}$	−456.5	−3.6
GEC1	5.0	$20_{(10)}$	53	$0.34_{(0.1)}$	17.6	3.5
GEC2	4.9	$24_{(9)}$	60	$0.57_{(0.2)}$	13.2	2.7
GEC3	4.7	$25_{(9)}$	61	$0.87_{(0.2)}$	8.9	1.9
NORM0.1	4.4	$9_{(1)}$	16	$0.46_{(0.1)}$	15.0	3.4
NORM0.5	5.0	$20_{(5)}$	41	$1.33_{(0.3)}$	11.6	2.3
NORM1	6.1	$28_{(8)}$	61	$2.14_{(0.7)}$	8.8	1.4
TARVM	4.6	$5_{(2)}$	15	$0.09_{(0.0)}$	21.9	4.8
TS-POET	4.2	$19_{(5)}$	37	$0.87_{(0.3)}$	17.2	4.1
PR-NERIVE	3.8	$11_{(3)}$	22	$0.43_{(0.1)}$	15.9	4.1
PR-quarNERIVE	3.9	$11_{(3)}$	23	$0.40_{(0.1)}$	16.1	4.2
PR-POET	4.0	$16_{(5)}$	33	$0.56_{(0.2)}$	17.0	4.3
RPR-POET	5.9	$15_{(4)}$	31	$0.64_{(0.2)}$	23.4	4.0
daily rebalancing portfolio with 1-day training window						
quarNERIVE	4.4	$8_{(3)}$	18	$0.77_{(0.2)}$	16.3	3.7
EQUAL	5.4	$1_{(-)}$	1	–	22.4	4.2
TSCV	85.5	$364_{(837)}$	7965	$42.63_{(846.9)}$	268.1	3.1
GEC1	5.2	$8_{(12)}$	69	$0.18_{(0.2)}$	20.4	3.9
GEC2	5.8	$20_{(13)}$	74	$0.86_{(0.2)}$	26.9	4.7
GEC3	10.2	$33_{(28)}$	157	$1.36_{(3.6)}$	3.0	0.3
NORM0.1	4.7	$8_{(2)}$	17	$0.78_{(0.2)}$	21.9	4.7
NORM0.5	8.6	$20_{(10)}$	51	$2.11_{(0.9)}$	29.2	3.4
NORM1	11.5	$28_{(16)}$	74	$3.20_{(2.3)}$	30.9	2.7
TARVM	332.8	$486_{(2383)}$	29145	$26.54_{(238.0)}$	684.5	2.1
TS-POET	5.4	$23_{(10)}$	82	$1.79_{(0.5)}$	20.7	3.9
PR-quarNERIVE	4.3	$8_{(2)}$	15	$0.81_{(0.2)}$	18.4	4.3
PR-POET	4.2	$13_{(5)}$	44	$0.96_{(0.3)}$	20.1	4.8
RPR-POET	6.5	$17_{(6)}$	49	$1.11_{(0.2)}$	21.0	3.2

Table 7: Empirical results (jumps removed) for the 73 most actively traded stocks in NYSE: annualized out-of-sample standard deviation, averaged maximum absolute weight, maximum of maximum absolute weight, annualized portfolio return and Sharpe ratio for various methods, including GEC with $c = 1, 2, 3$, and NORM with $\delta = 0.1, 0.5, 1$.

asset, which should satisfy

$$v_{s-1} < t_s^{(i_1)} \leq t_s^{(i_2)} \leq \cdots \leq t_s^{(i_p)} = v_s,$$

where $\{i_1, \ldots, i_p\}$ is some permutation of $1, \ldots, p$. Letting b_s denote the number of tides, we can write the above as

$$v_{s-1} < t_s^{j_1} < t_s^{j_2} < \cdots < t_s^{j_{p-b_s}} = v_s,$$

where $j_1, \ldots, j_{p-b_s} \in \{1, \ldots, p\}$.

Then we can write, for $s = 1, \ldots, nL$,

$$\mathbf{X}_{v_s} - \mathbf{X}(s) = \sum_{m=1}^{p-b_s-1} \mathbf{D}_m^s \mathbf{A}(t_s^{j_m}, t_s^{j_{m+1}}) \mathbf{Z}_{d,s}^j(m+1) + \sum_{m=1}^{p-b_s-1} \mathbf{D}_m^s \mathbf{\Sigma}(t_s^{j_m}, t_s^{j_{m+1}})^{1/2} \mathbf{Z}_{v,s}^j(m+1), \tag{7.1}$$

where $\mathbf{D}_m^s$ is a diagonal matrix with either 0 or 1 as elements. The jth diagonal element is 1 if the jth asset is already traded at time $t_s^{j_m}$, and 0 otherwise. The matrices $\mathbf{A}(\cdot, \cdot)$ and $\mathbf{\Sigma}(\cdot, \cdot)$ are as defined in

Assumption (D1) and (V1) respectively.

(A3) If the drift $\boldsymbol{\mu}_t$ is random, the components of $\mathbf{Z}^j_{d,s}(m+1)$, $\mathbf{Z}^j_{v,s}(m+1) \in \mathcal{F}^j_{t^{j_{m+1}}_s}$ are conditionally independent given $\mathcal{F}_{-j}$, $E(\mathbf{Z}^j_{d,s}(m+1)|\mathcal{F}_{-j}) = \mathbf{0} = E(\mathbf{Z}^j_{v,s}(m+1)|\mathcal{F}_{-j})$, $\mathrm{var}(\mathbf{Z}^j_{d,s}(m+1)|\mathcal{F}_{-j}) = \mathbf{I}_p = \mathrm{var}(\mathbf{Z}^j_{v,s}(m+1)|\mathcal{F}_{-j})$ almost surely. Eighth moments exist for their components as well.

If the drift $\boldsymbol{\mu}_t$ is non-random, then $\mathbf{Z}^j_{d,s}(m+1) = (1, 0, \ldots, 0)^{\mathrm{T}}$.

(A4) (Only for random drift). Using notations in Assumption (D2), we assume that for some $c_{d,j,s} \in \mathcal{F}_{-j} \cup \mathcal{F}^j_s$ greater than 0, and for $\ell = 1, \ldots, m$,

$$
\begin{aligned}
&E\Big(\mathbf{p}^{\mathrm{T}}_{ij}\mathbf{D}^s_m\mathbf{A}(t^{j_m}_s, t^{j_{m+1}}_s)\mathbf{Z}^j_{d,s}(\ell+1)\big|\mathcal{F}_{-j} \cup \mathcal{F}^j_{t^{j_\ell}_s}\Big) \\
&= \left(1 - \frac{c_{d,j,s}}{(p - b_s - 1)^{1/6}}\right)\mathbf{p}^{\mathrm{T}}_{ij}\mathbf{D}^s_m\mathbf{A}(t^{j_m}_s, t^{j_{m+1}}_s)\mathbf{Z}^j_{d,s}(\ell) + e^{ij}_{d,s}(\ell),
\end{aligned}
$$

where we define $\mathbf{Z}^j_{d,s}(\ell)\mathbf{Z}^j_{d,s}(\ell)^{\mathrm{T}} = \mathbf{I}_p$ and $e^{ij}_{d,s}(\ell) = 0$ for $\ell \leq 0$. The process $\{e^{ij}_{d,s}(\ell)\}$ with $e^{ij}_{d,s}(\ell) \in \mathcal{F}^j_{t^{j_\ell}_s}$ has $E(e^{ij}_{d,s}(\ell)|\mathcal{F}_{-j} \cup \mathcal{F}^j_{t^{j_{\ell-1}}_s}) = 0$ almost surely, and $e^{ij}_{d,s}(\ell)|\mathcal{F}_{-j} \cup \mathcal{F}^j_{t^{j_{\ell-1}}_s} = O_P(\big\|\mathbf{A}(t^{j_{\ell-1}}_s, t^{j_\ell}_s)\big\|) = O_P(p^{1/2} \cdot (p - b_s - 1)^{-1}n^{-1}L^{-1})$.

The assumption for $E\Big(\mathbf{p}^{\mathrm{T}}_{ij}\mathbf{D}^s_m\boldsymbol{\Sigma}(t^{j_m}_s, t^{j_{m+1}}_s)^{1/2}\mathbf{Z}^j_{v,s}(\ell+1)\big|\mathcal{F}_{-j} \cup \mathcal{F}^j_{t^{j_\ell}_s} \cup \mathcal{F}^{\boldsymbol{\sigma}}_{v_s}\Big)$ runs parallel to the above, with $c_{v,j,s} \in \mathcal{F}_{-j} \cup \mathcal{F}^j_s$ replaces $c_{d,j,s}$, $\mathbf{Z}^j_{v,s}(\cdot)$ replaces $\mathbf{Z}^j_{d,s}(\cdot)$, and $e^{ij}_{v,s}(\cdot)$ replaces $e^{ij}_{d,s}(\cdot)$ with

$$
e^{ij}_{v,s}(\ell)|\mathcal{F}_{-j} \cup \mathcal{F}^j_{t^{j_{\ell-1}}_s} \cup \mathcal{F}^{\boldsymbol{\sigma}}_{v_s} = O_P(\big\|\boldsymbol{\Sigma}(t^{j_{\ell-1}}_s, t^{j_\ell}_s)^{1/2}\big\|) = O_P(\big\|\mathbf{A}(t^{j_{\ell-1}}_s, t^{j_\ell}_s)\big\|/|t^{j_\ell}_s - t^{j_{\ell-1}}_s|^{1/2}).
$$

(A5) (Only for random drift). Let $\psi(x) = e^{x^2} - 1$. We assume that for $\ell = 1, \ldots, m$,

$$
\begin{aligned}
&E\left\{\psi\left(\frac{|\mathbf{p}^{\mathrm{T}}_{ij}\mathbf{D}^s_m\mathbf{A}(t^{j_m}_s, t^{j_{m+1}}_s)\mathbf{Z}^j_{d,s}(\ell)|}{|\mathbf{p}^{\mathrm{T}}_{ij}\mathbf{D}^s_m\mathbf{A}(t^{j_m}_s, t^{j_{m+1}}_s)\mathbf{Z}^j_{d,s}(\ell-1)|}\right)\Big|\mathcal{F}_{-j} \cup \mathcal{F}^j_{t^{j_{\ell-1}}_s}\right\} < \infty, \\
&E\left\{\psi\left(\frac{|e^{ij}_{d,s}(\ell)|}{|\mathbf{p}^{\mathrm{T}}_{ij}\mathbf{D}^s_m\mathbf{A}(t^{j_m}_s, t^{j_{m+1}}_s)\mathbf{Z}^j_{d,s}(\ell-1)|}\right)\Big|\mathcal{F}_{-j} \cup \mathcal{F}^j_{t^{j_{\ell-1}}_s}\right\} < \infty.
\end{aligned}
$$

The assumption for the volatility runs parallel to the above, with the expectations now conditional on $\mathcal{F}_{-j} \cup \mathcal{F}^j_{t^{j_{\ell-1}}_s} \cup \mathcal{F}^{\boldsymbol{\sigma}}_{v_s}$, $\boldsymbol{\Sigma}(\cdot,\cdot)^{1/2}$ replaces $\mathbf{A}(\cdot,\cdot)$, $\mathbf{Z}^j_{v,s}(\cdot)$ replaces $\mathbf{Z}^j_{d,s}(\cdot)$ and $e^{ij}_{v,s}(\cdot)$ replaces $e^{ij}_{d,s}(\cdot)$.

Assumptions (A3), (A4) and (A5) are parallel to (D1), (D2) and (D3) respectively. The major difference is that the coefficients $\rho^j_{d,K,q}, \rho^j_{v,K,q} \leq \xi < 1$ are now replaced by coefficients that are going to 1 as $n, p \to \infty$. This reflects that the correlations among variables between tick-by-tick trading times are high, since the time length between ticks is usually very small. Note that if the drift is non-random, we only need Assumption (A3) that $\mathbf{Z}^j_{d,s}(m+1) = (1, 0, \ldots, 0)^{\mathrm{T}}$, which is just a matter of notation rather than an assumption.

We provide the proof of all the theorems of the paper in this section. We assume the jump-diffusion model (3.2) for the log-price process $\{\mathbf{X}_t\}$, and prove Theorem 4, so that Theorem 2 then follows automatically. Define

$$\widetilde{\mathbf{Y}}_t = \mathbf{Y}_t - \widehat{\mathbf{J}}_t = (\mathbf{X}_t - \widehat{\mathbf{J}}_t) + \boldsymbol{\epsilon}_t = \widetilde{\mathbf{X}}_t + \boldsymbol{\epsilon}_t, \tag{7.2}$$

where $\{\widehat{\mathbf{J}}_t\}$ is the estimated jump process using the wavelet method in Fan and Wang (2007) described in Section 3.1. Then $\{\widetilde{\mathbf{X}}_t\}$ represents the jumps-removed log-price process. For $j = 1, \ldots, L$ and $v_s = v_s^j$ for $s = 0, \ldots, n(j)$, we then have

$$\widetilde{\mathbf{Y}}(s) = \widetilde{\mathbf{X}}(s) + \boldsymbol{\epsilon}(s) = \widetilde{\mathbf{X}}_{v_s} + \mathbf{E}(s),$$

where we define

$$\mathbf{E}(s) = \boldsymbol{\epsilon}(s) + \widetilde{\mathbf{X}}(s) - \widetilde{\mathbf{X}}_{v_s} = \boldsymbol{\epsilon}(s) + (\mathbf{X}(s) - \widehat{\mathbf{J}}(s)) - (\mathbf{X}_{v_s} - \widehat{\mathbf{J}}_{v_s}).$$

We can then decompose, for $i = 1, \ldots, p$, $j = 1, \ldots, L$ with $\mathbf{P}_{-j} = (\mathbf{p}_{1j}, \ldots, \mathbf{p}_{pj})$,

$$\begin{aligned}
\mathbf{p}_{ij}^{\mathrm{T}} \widetilde{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij} &= \mathbf{p}_{ij}^{\mathrm{T}} [\widetilde{\mathbf{Y}}, \widetilde{\mathbf{Y}}^{\mathrm{T}}]_j^{(K)} \mathbf{p}_{ij} - \frac{|S^j(K)|_K}{|S^j(1)|} \mathbf{p}_{ij}^{\mathrm{T}} [\widetilde{\mathbf{Y}}, \widetilde{\mathbf{Y}}^{\mathrm{T}}]_j^{(1)} \mathbf{p}_{ij} \\
&= I_1 + 2I_2 + I_3,
\end{aligned}$$

where $\widetilde{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j)$ is the TSCV in (2.4) constructed using jumps-removed data, and

$$\begin{aligned}
I_1 &= \mathbf{p}_{ij}^{\mathrm{T}} [\widetilde{\mathbf{X}}_v, \widetilde{\mathbf{X}}_v^{\mathrm{T}}]_j^{(K)} \mathbf{p}_{ij} - \frac{|S^j(K)|_K}{|S^j(1)|} \mathbf{p}_{ij}^{\mathrm{T}} [\widetilde{\mathbf{X}}_v, \widetilde{\mathbf{X}}_v^{\mathrm{T}}]_j^{(1)} \mathbf{p}_{ij}, \\
I_2 &= \mathbf{p}_{ij}^{\mathrm{T}} [\widetilde{\mathbf{X}}_v, \mathbf{E}^{\mathrm{T}}]_j^{(K)} \mathbf{p}_{ij} - \frac{|S^j(K)|_K}{|S^j(1)|} \mathbf{p}_{ij}^{\mathrm{T}} [\widetilde{\mathbf{X}}_v, \mathbf{E}^{\mathrm{T}}]_j^{(1)} \mathbf{p}_{ij}, \\
I_3 &= \mathbf{p}_{ij}^{\mathrm{T}} [\mathbf{E}, \mathbf{E}^{\mathrm{T}}]_j^{(K)} \mathbf{p}_{ij} - \frac{|S^j(K)|_K}{|S^j(1)|} \mathbf{p}_{ij}^{\mathrm{T}} [\mathbf{E}, \mathbf{E}^{\mathrm{T}}]_j^{(1)} \mathbf{p}_{ij},
\end{aligned} \tag{7.3}$$

with $[\widetilde{\mathbf{X}}_v, \widetilde{\mathbf{X}}_v^{\mathrm{T}}]_j^{(m)}$, $[\widetilde{\mathbf{X}}_v, \mathbf{E}^{\mathrm{T}}]_j^{(m)}$ and $[\mathbf{E}, \mathbf{E}^{\mathrm{T}}]_j^{(m)}$ defined by

$$\begin{aligned}
[\widetilde{\mathbf{X}}_v, \widetilde{\mathbf{X}}_v^{\mathrm{T}}]_j^{(m)} &= \frac{1}{m} \sum_{s, s+m \in S^j(m)} (\widetilde{\mathbf{X}}_{v_{s+m}} - \widetilde{\mathbf{X}}_{v_s})(\widetilde{\mathbf{X}}_{v_{s+m}} - \widetilde{\mathbf{X}}_{v_s})^{\mathrm{T}}, \\
[\widetilde{\mathbf{X}}_v, \mathbf{E}^{\mathrm{T}}]_j^{(m)} &= \frac{1}{m} \sum_{s, s+m \in S^j(m)} (\widetilde{\mathbf{X}}_{v_{s+m}} - \widetilde{\mathbf{X}}_{v_s})(\mathbf{E}(s+m) - \mathbf{E}(s))^{\mathrm{T}}, \\
[\mathbf{E}, \mathbf{E}^{\mathrm{T}}]_j^{(m)} &= \frac{1}{m} \sum_{s, s+m \in S^j(m)} (\mathbf{E}(s+m) - \mathbf{E}(s))(\mathbf{E}(s+m) - \mathbf{E}(s))^{\mathrm{T}}.
\end{aligned}$$

Lemma 1. *Let all the assumptions in Theorem 4 hold. Then with $p/n \to c > 0$ when there are no pervasive*

factors, or $p^{3/2}/n \to c > 0$ *when there are pervasive factors,*

$$\max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}} \left| \frac{I_1}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} - 1 \right| = O_P(n^{-1/6}).$$

Proof of Lemma 1. By Assumption (D1) and (V1), we first decompose for an integer $m \geq 1$, and $i = 1, \ldots, p$, $j = 1, \ldots, L$,

$$\begin{aligned}
\mathbf{p}_{ij}^{\mathrm{T}} [\widetilde{\mathbf{X}}_v, \widetilde{\mathbf{X}}_v^{\mathrm{T}}]_j^{(m)} \mathbf{p}_{ij} &= I_{11} + 2I_{12} + I_{13}, \quad \text{where} \\
I_{11} &= \frac{1}{m} \sum_{q=0}^{m-1} \sum_{r=1}^{|S^j(m)|_m} (\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{A}(v_{(r-1)m+q}, v_{rm+q}) \mathbf{Z}_{d,rm+q}^j + \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(v_{(r-1)m+q}, v_{rm+q})^{1/2} \mathbf{Z}_{v,rm+q}^j)^2, \\
I_{12} &= \frac{1}{m} \sum_{q=0}^{m-1} \sum_{r=1}^{|S^j(m)|_m} (\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{A}(v_{(r-1)m+q}, v_{rm+q}) \mathbf{Z}_{d,rm+q}^j + \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(v_{(r-1)m+q}, v_{rm+q})^{1/2} \mathbf{Z}_{v,rm+q}^j) \\
&\qquad \cdot (\mathbf{J}_{v_{rm+q}} - \widehat{\mathbf{J}}_{v_{rm+q}} - \mathbf{J}_{v_{(r-1)m+q}} + \widehat{\mathbf{J}}_{v_{(r-1)m+q}})^{\mathrm{T}} \mathbf{p}_{ij}, \\
I_{13} &= \frac{1}{m} \sum_{q=0}^{m-1} \sum_{r=1}^{|S^j(m)|_m} \left((\mathbf{J}_{v_{rm+q}} - \widehat{\mathbf{J}}_{v_{rm+q}} - \mathbf{J}_{v_{(r-1)m+q}} + \widehat{\mathbf{J}}_{v_{(r-1)m+q}})^{\mathrm{T}} \mathbf{p}_{ij} \right)^2.
\end{aligned} \tag{7.4}$$

Consider further decomposition

$$\begin{aligned}
\left| \frac{I_{11}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} - 1 \right| &\leq \left| \frac{1}{m} \sum_{q=0}^{m-1} \sum_{r=1}^{|S^j(m)|_m} (a_{d,r,m,q}^{ij}(r))^2 \right| + \left| \frac{2}{m} \sum_{q=0}^{m-1} \sum_{r=1}^{|S^j(m)|_m} a_{d,r,m,q}^{ij}(r) b_{v,r,m,q}^{ij}(r) \right| \\
&\quad + \left| \frac{1}{m} \sum_{q=0}^{m-1} \sum_{r=1}^{|S^j(m)|_m} (b_{v,r,m,q}^{ij}(r))^2 - 1 \right|, \text{ where} \\
(a_{d,r,m,q}^{ij}(\ell))^2 &= (\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{A}(v_{(r-1)m+q}, v_{rm+q}) \mathbf{Z}_{d,\ell m+q}^j)^2 / \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}, \\
(b_{v,r,m,q}^{ij}(\ell))^2 &= (\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(v_{(r-1)m+q}, v_{rm+q})^{1/2} \mathbf{Z}_{v,\ell m+q}^j)^2 / \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}.
\end{aligned}$$

To find the order of $I_{11}/\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij} - 1$, define

$$\begin{aligned}
g_{d,r,m,q}^{ij}(\ell) &= (a_{d,r,m,q}^{ij}(\ell))^2 - E((a_{d,r,m,q}^{ij}(\ell))^2 | \mathcal{F}_{-j} \cup \mathcal{F}_{(\ell-1)m+q}^j), \\
g_{v,r,m,q}^{ij}(\ell) &= (b_{v,r,m,q}^{ij}(\ell))^2 - E((b_{v,r,m,q}^{ij}(\ell))^2 | \mathcal{F}_{-j} \cup \mathcal{F}_{(\ell-1)m+q}^j \cup \mathcal{F}_{v_{rm+q}}^{\boldsymbol{\sigma}}).
\end{aligned}$$

Then we first consider

$$
\begin{aligned}
&\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}(a^{ij}_{d,r,m,q}(r))^2\\
&=\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}[(a^{ij}_{d,r,m,q}(r))^2-E((a^{ij}_{d,r,m,q}(r))^2|\mathcal{F}_{-j}\cup\mathcal{F}^j_{(r-1)m+q})]\\
&+\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}\Bigg\{\rho^j_{d,m,q}(a^{ij}_{d,r,m,q}(r-1))^2\\
&+(1-\rho^j_{d,m,q})\frac{\left\|\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{(r-1)m+q},v_{rm+q})\right\|^2}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}+\frac{e^{ij}_{d,(r-1)m+q}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\Bigg\}\\
&=\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}g^{ij}_{d,r,m,q}(r)+\rho^j_{d,m,q}\cdot\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=2}^{|S^j(m)|_m}g^{ij}_{d,r,m,q}(r-1)\\
&+\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}\left\{\frac{e^{ij}_{d,(r-1)m+q}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}+\rho^j_{d,m,q}\cdot\frac{e^{ij}_{d,(r-2)m+q}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\right\}\\
&+(\rho^j_{d,m,q})^2\cdot\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=3}^{|S^j(m)|_m}\left\{(a^{(ij)}_{d,r,m,q}(r-2))^2-\frac{\left\|\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{(r-1)m+q},v_{rm+q})\right\|^2}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\right\}\\
&+\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}\frac{\left\|\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{(r-1)m+q},v_{rm+q})\right\|^2}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\\
&=I_{11,1}+I_{11,2}+I_{11,3}+I_{11,4},
\end{aligned}
$$

where the equalities use Assumption (D2), and

$$
\begin{aligned}
I_{11,1}&=\frac{1}{m}\sum_{q=0}^{m-1}\left\{\sum_{\ell=0}^{\lfloor|S^j(m)|_m/2\rfloor-1}(\rho^j_{d,m,q})^\ell\sum_{r=1+\ell}^{|S^j(m)|_m}g^{ij}_{d,r,m,q}(r-\ell)\right\},\\
I_{11,2}&=\frac{1}{m}\sum_{q=0}^{m-1}\left\{\sum_{\ell=0}^{\lfloor|S^j(m)|_m/2\rfloor-1}(\rho^j_{d,m,q})^\ell\sum_{r=1+\ell}^{|S^j(m)|_m}\frac{e^{ij}_{d,(r-1-\ell)m+q}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\right\},\\
I_{11,3}&=(\rho^j_{d,m,q})^{\lfloor|S^j(m)|_m/2\rfloor}\\
&\quad\cdot\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=\lfloor|S^j(m)|_m/2\rfloor+1}^{|S^j(m)|_m}\left\{(a^{(ij)}_{d,r,m,q}(r-2))^2-\frac{(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{(r-1)m+q},v_{rm+q}))^2}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\right\},\\
I_{11,4}&=\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}\frac{(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{(r-1)m+q},v_{rm+q}))^2}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}.
\end{aligned}
$$

Letting $K^{ij}_{d,r,m,q}(\ell) = \frac{(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{(r-1)m+q}, v_{rm+q})\mathbf{Z}^j_{d,(\ell-1)m+q})^2}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}$, by Assumption (D2),

$$\begin{aligned} &E\left\{\psi\left(\frac{|g^{ij}_{d,r,m,q}(r-\ell)|}{K^{ij}_{d,r,m,q}(r-\ell)}\right)\Big|\mathcal{F}_{-j}\cup\mathcal{F}^j_{(r-1-\ell)m+q}\cup\mathcal{F}^{\boldsymbol{\sigma}}_{\tau_j}\right\}<\infty,\\ &E\left\{\psi\left(\frac{|e^{ij}_{d,(r-1-\ell)m+q}/\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}|}{K^{ij}_{d,r,m,q}(r-1-\ell)}\right)\Big|\mathcal{F}_{-j}\cup\mathcal{F}^j_{(r-2-\ell)m+q}\cup\mathcal{F}^{\boldsymbol{\sigma}}_{\tau_j}\right\}<\infty. \end{aligned} \tag{7.5}$$

At the same time, by Assumption (D1) that eighth moments exist for the $\mathbf{Z}^j_{d,(r-1-\ell)m+q}$'s and are conditionally independent given $\mathcal{F}_{-j}$, we can use Lemma 2.7 of Bai and Silverstein (1998) to arrive at

$$\begin{aligned} E((K^{ij}_{d,r,m,q}(r-\ell))^4|\mathcal{F}_{-j}\cup\mathcal{F}^{\boldsymbol{\sigma}}_{\tau_j}) &= O(\big\|\mathbf{A}(v_{(r-1)m+q}, v_{rm+q})\big\|^8/(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij})^4)\\ &= O\Big(p_f m\cdot\frac{1}{(nL)^2}/\frac{1}{L}\Big)^4 = O\Big(\frac{p_f m}{n^2L}\Big)^4, \text{ so that}\\ K^{ij}_{d,r,m,q}(r-\ell)^2 &= O_P\Big(p_f m\cdot\frac{1}{(nL)^2}/\frac{1}{L}\Big)^2 = O_P\Big(\frac{p_f m}{n^2L}\Big)^2, \end{aligned} \tag{7.6}$$

where the last line used Assumption (D1), with $p_f = 1$ if there are no pervasive factors and $p_f = p$ if there are pervasive factors or the drift is non-random, and the second line used Assumption (V1) on the rate of $\lambda_{\min}(\mathbf{\Sigma}(\tau_{j-1},\tau_j))$. With (7.5) and (7.6), we can apply Theorem 2.2 of van de Geer (2002) to arrive at

$$\sum_{r=1+\ell}^{|S^j(m)|_m} g^{ij}_{d,r,m,q}(r-\ell),\ \sum_{r=1+\ell}^{|S^j(m)|_m}\frac{e^{ij}_{d,(r-1-\ell)m+q}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}} = O_P\Big(|S^j(m)|_m^{1/2}\cdot\frac{p_f m}{n^2L}\Big) = O_P\Big(\frac{p_f m^{1/2}}{n^{3/2}L}\Big),$$

for $\ell = 0, 1, \ldots, \lfloor |S^j(m)|_m/2\rfloor - 1$. Since $\rho^j_{d,m,q} \le \xi < 1$ uniformly by Assumption (D2), we have

$$I_{11,1},\ I_{11,2} = O_P\left(\frac{p_f m^{1/2}}{n^{3/2}L}\right). \tag{7.7}$$

Similar techniques in finding the order of $K^{ij}_{d,r,m,q}(r-\ell)$ show that

$$I_{11,3} = O_P\Big(\xi^{n/m}\cdot\frac{p_f m}{n^2L}\Big). \tag{7.8}$$

For $I_{11,4}$, by (7.6), we have

$$I_{11,4} = O_P\Big(|S^j(m)|_m\cdot\frac{p_f m}{n^2L}\Big) = O_P\Big(\frac{p_f}{nL}\Big). \tag{7.9}$$

Combining (7.7), (7.8) and (7.9), we have

$$\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}(a^{ij}_{d,r,m,q}(r))^2 = O_P(p_f n^{-1}L^{-1}). \tag{7.10}$$

Similar to the above calculations, by Assumption (V2), we can decompose

$$\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}(b^{ij}_{v,r,m,q}(r))^2-1=J_{11,1}+J_{11,2}+J_{11,3}+J_{11,4},\text{ where}$$

$$J_{11,1}=\frac{1}{m}\sum_{q=0}^{m-1}\left\{\sum_{\ell=0}^{\lfloor|S^j(m)|_m/2\rfloor-1}(\rho^j_{v,m,q})^\ell\sum_{r=1+\ell}^{|S^j(m)|_m}g^{ij}_{v,r,m,q}(r-\ell)\right\},$$

$$J_{11,2}=\frac{1}{m}\sum_{q=0}^{m-1}\left\{\sum_{\ell=0}^{\lfloor|S^j(m)|_m/2\rfloor-1}(\rho^j_{v,m,q})^\ell\sum_{r=1+\ell}^{|S^j(m)|_m}\frac{e^{ij}_{v,(r-1-\ell)m+q}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\right\},$$

$$J_{11,3}=(\rho^j_{v,m,q})^{\lfloor|S^j(m)|_m/2\rfloor}\cdot\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=\lfloor|S^j(m)|_m/2\rfloor+1}^{|S^j(m)|_m}\left\{(b^{(ij)}_{v,r,m,q}(r-2))^2-\frac{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(v_{(r-1)m+q},v_{rm+q})\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\right\},$$

$$J_{11,4}=\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m}\frac{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(v_{(r-1)m+q},v_{rm+q})\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}-1.$$

Letting $K^{ij}_{v,r,m,q}(\ell)=\frac{(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(v_{(r-1)m+q},v_{rm+q})^{1/2}\mathbf{Z}^j_{d,(\ell-1)m+q})^2}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}$, by Assumption (V2),

$$\begin{aligned}&E\left\{\psi\left(\frac{|g^{ij}_{v,r,m,q}(r-\ell)|}{K^{ij}_{v,r,m,q}(r-\ell)}\right)\Big|\mathcal{F}_{-j}\cup\mathcal{F}^j_{(r-1-\ell)m+q}\right\}<\infty,\\&E\left\{\psi\left(\frac{|e^{ij}_{v,(r-1-\ell)m+q}/\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}|}{K^{ij}_{v,r,m,q}(r-1-\ell)}\right)\Big|\mathcal{F}_{-j}\cup\mathcal{F}^j_{(r-2-\ell)m+q}\right\}<\infty.\end{aligned}\tag{7.11}$$

At the same time, by Assumption (V1) that eighth moments exist for the $\mathbf{Z}^j_{v,(r-1-\ell)m+q}$'s and are conditionally independent given $\mathcal{F}_{-j}$, we can use Lemma 2.7 of Bai and Silverstein (1998) to arrive at

$$\begin{aligned}E((K^{ij}_{v,r,m,q}(r-\ell))^4|\mathcal{F}_{-j}\cup\mathcal{F}^{\boldsymbol{\sigma}}_{\tau_j})&=O((\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(v_{(r-1)m+q},v_{rm+q})\mathbf{p}_{ij})^4/(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij})^4)\\&=O\Big(p_f\cdot\frac{m}{nL}/\frac{p_f}{L}\Big)^4,\text{ so that}\\K^{ij}_{v,r,m,q}(r-\ell)^2&=O_P\Big(p_f\cdot\frac{m}{nL}/\frac{p_f}{L}\Big)^2=O_P\Big(\frac{m}{n}\Big)^2,\end{aligned}\tag{7.12}$$

where the last line used Assumption (V1), with $p_f=1$ if there are no pervasive factors and $p_f=p$ if there are pervasive factors. The main difference between (7.6) and (7.12) is that in (7.12), the numerator is a part of the denominator, and if pervasive factors affect the numerator, they have to affect the denominator too. This results in the balance of orders and hence p_f disappears from the order of the term. With (7.11) and (7.12), we can apply Theorem 2.2 of van de Geer (2002) to arrive at

$$J_{11,1},\ J_{11,2}=O_P\left(|S^j(m)|_m^{1/2}\cdot\frac{m}{n}\right)=O_P(m^{1/2}n^{-1/2}).\tag{7.13}$$

Similar to $I_{11,3}$, we have

$$J_{11,3} = O_P(\xi^{n/m} \cdot mn^{-1}). \tag{7.14}$$

For $J_{11,4}$, using Assumption (V1),

$$\begin{aligned} J_{11,4} &= \frac{1}{m}\sum_{q=0}^{m-1} \frac{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(v_q, v_{n(j)-m+1+q})\mathbf{p}_{ij} - \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}} \\ &= -\frac{1}{m}\sum_{q=0}^{m-1} \frac{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(v_{n(j)-m+1+q}, \tau_j)\mathbf{p}_{ij} + \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1}, v_q)\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}} \\ &= O_P\left(\frac{1}{m}\sum_{q=0}^{m-1} \frac{(m-1-q)+q}{nL} / \frac{1}{L}\right) = O_P(mn^{-1}). \end{aligned} \tag{7.15}$$

Combining (7.13), (7.14) and (7.15), we have

$$\frac{1}{m}\sum_{q=0}^{m-1}\sum_{r=1}^{|S^j(m)|_m} (b_{v,r,m,q}^{ij}(r))^2 - 1 = O_P(m^{1/2}n^{-1/2}). \tag{7.16}$$

Using the Cauchy-Schwarz inequality, together with (7.10) and (7.16), we have

$$\begin{aligned} \left|\frac{I_{11}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}} - 1\right| &= O_P(p_f n^{-1}L^{-1} + m^{1/2}n^{-1/2} + p_f^{1/2}n^{-1/2}L^{-1/2}) \\ &= O_P(n^{-1/6}), \end{aligned} \tag{7.17}$$

if there are pervasive factors such that $p_f = p \asymp n^{2/3}$ and $m = O(n^{2/3})$. Turning to I_{12} and I_{13} defined in (7.3), using Assumption (W1) to (W3), and the rate in Fan and Wang (2007), we have

$$I_{13}/\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij} = O_P(n^{-1/2}L^{1/2}).$$

The above implies, through using the Cauchy-Schwarz inequality,

$$I_{12}/\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij} = O_P(n^{-1/4}L^{1/4}).$$

Combining all results, we have for $K \asymp n^{2/3}$,

$$\begin{aligned} \left|\frac{\mathbf{p}_{ij}^{\mathrm{T}}[\widetilde{\mathbf{X}}_v, \widetilde{\mathbf{X}}_v^{\mathrm{T}}]_j^{(K)}\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}} - 1\right| &= O_P(n^{-1/6}), \\ \frac{|S^j(K)|_K}{|S^j(1)|} \cdot \left|\frac{\mathbf{p}_{ij}^{\mathrm{T}}[\widetilde{\mathbf{X}}_v, \widetilde{\mathbf{X}}_v^{\mathrm{T}}]_j^{(1)}\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\right| &= O_p(K^{-1}\cdot 1) = O_P(n^{-2/3}). \end{aligned}$$

Note that the above bounds are independent of the indices i and j, and hence

$$\max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}} \left| \frac{I_1}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} - 1 \right| = O_P(n^{-1/6} + n^{-2/3}) = O_P(n^{-1/6}).$$

This completes of proof of the lemma. □

Lemma 2. *Let all the assumptions in Theorem 4 hold. Then with $p/n \to c > 0$ when there are no pervasive factors, or $p^{3/2}/n \to c > 0$ when there are pervasive factors,*

$$\max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}} \max_{s=1,\ldots,n(j)} \left| \frac{\mathbf{p}_{ij}^{\mathrm{T}} (\mathbf{X}_{v_s} - \mathbf{X}(s))}{(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}} \right| = O_P(p^{1/6} n^{-1/2}).$$

Proof of Lemma 2. Consider $\frac{\mathbf{p}_{ij}^{\mathrm{T}} (\mathbf{X}_{v_s} - \mathbf{X}(s))}{(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}} = A_d^{ij}(s) + A_v^{ij}(s)$, where using (7.1),

$$A_d^{ij}(s) = \sum_{m=1}^{p-b_s-1} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{A}(t_s^{j_m}, t_s^{j_{m+1}}) \mathbf{Z}_{d,s}^j(m+1)}{(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}},$$
$$A_v^{ij}(s) = \sum_{m=1}^{p-b_s-1} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{\Sigma}(t_s^{j_m}, t_s^{j_{m+1}})^{1/2} \mathbf{Z}_{v,s}^j(m+1)}{(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}}.$$

We first deal with non-random drift for $A_d^{ij}(s)$. By Assumptions (D1) and (V1), we have

$$|A_d^{ij}(s)| \le \sum_{m=1}^{p-b_s-1} \frac{\left\| \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{A}(t_s^{j_m}, t_s^{j_{m+1}}) \right\|}{(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}} = O_P((p - b_s - 1) \cdot p^{1/2} \cdot (p - b_s - 1)^{-1} n^{-1} L^{-1} / L^{1/2})$$
$$= O_P(p^{1/2} n^{-1}). \tag{7.18}$$

Now we focus on random drift. Define for $\ell = 1, \ldots, m+1$,

$$g_{d,s,m}^{ij}(\ell) = \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{A}(t_s^{j_m}, t_s^{j_{m+1}}) \mathbf{Z}_{d,s}^j(\ell) - E(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{A}(t_s^{j_m}, t_s^{j_{m+1}}) \mathbf{Z}_{d,s}^j(\ell) | \mathcal{F}_{-j} \cup \mathcal{F}_{t_s^{j_{\ell-1}}}^j)}{(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}},$$
$$g_{v,s,m}^{ij}(\ell) = \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{\Sigma}(t_s^{j_m}, t_s^{j_{m+1}})^{1/2} \mathbf{Z}_{v,s}^j(\ell) - E(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{\Sigma}(t_s^{j_m}, t_s^{j_{m+1}})^{1/2} \mathbf{Z}_{v,s}^j(\ell) | \mathcal{F}_{-j} \cup \mathcal{F}_{t_s^{j_{\ell-1}}}^j \cup \mathcal{F}_{\tau_j}^{\boldsymbol{\sigma}})}{(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}}.$$

Consider $A_d^{ij}(s)$ first. By Assumption (A4), we can decompose

$$\begin{aligned} A_d^{ij}(s) &= \sum_{m=1}^{p-b_s-1} g_{d,s,m}^{ij}(m+1) + \left(1 - \frac{c_{d,j,s}}{(p-b_s-1)^{1/6}}\right) \sum_{m=1}^{p-b_s-1} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{A}(t_s^{j_m}, t_s^{j_{m+1}}) \mathbf{Z}_{d,s}^j(m)}{(\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}} \\ &\quad + \sum_{m=1}^{p-b_s-1} \frac{e_{d,s}^{ij}(m)}{(\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}} \\ &= J_1 + J_2 + J_3, \end{aligned}$$

where

$$\begin{aligned} J_1 &= \sum_{\ell=0}^{\lfloor (p-b_s-1)/2 \rfloor - 1} \left(1 - \frac{c_{d,j,s}}{(p-b_s-1)^{1/6}}\right)^{\ell} \sum_{m=1}^{p-b_s-1} g_{d,s,m}^{ij}(m-\ell+1), \\ J_2 &= \sum_{\ell=0}^{\lfloor (p-b_s-1)/2 \rfloor - 1} \left(1 - \frac{c_{d,j,s}}{(p-b_s-1)^{1/6}}\right)^{\ell} \sum_{m=1}^{p-b_s-1} \frac{e_{d,s}^{ij}(m-\ell)}{(\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}}, \\ J_3 &= \left(1 - \frac{c_{d,j,s}}{(p-b_s-1)^{1/6}}\right)^{\lfloor (p-b_s-1)/2 \rfloor} \sum_{m=1}^{p-b_s-1} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{A}(t_s^{j_m}, t_s^{j_{m+1}}) \mathbf{Z}_{d,s}^j(m+1-\lfloor (p-b_s-1)/2 \rfloor)}{(\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}}. \end{aligned}$$

Letting $K_{d,s,m}^{ij}(\ell) = \frac{|\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{D}_m^s \mathbf{A}(t_s^{j_m}, t_s^{j_{m+1}}) \mathbf{Z}_{d,s}^j(\ell-1)|}{(\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}}$, by Assumption (A5),

$$\begin{aligned} &E\left\{\psi\left(\frac{|g_{d,s,m}^{ij}(m-\ell+1)|}{K_{d,s,m}^{ij}(m-\ell+1)}\right) \Big| \mathcal{F}_{-j} \cup \mathcal{F}_{t_s^{j_{m-\ell}}}^j \right\} < \infty, \\ &E\left\{\psi\left(\frac{|e_{d,(s}^{ij}(m-\ell)/(\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^{1/2}|}{K_{d,s,m}^{ij}(m-\ell)}\right) \Big| \mathcal{F}_{-j} \cup \mathcal{F}_{t_s^{j_{m-\ell-1}}}^j \right\} < \infty. \end{aligned} \tag{7.19}$$

At the same time, by Assumption (A3) that fourth moments exist for the $\mathbf{Z}_{d,s}^j(\ell)$'s and are conditionally independent given $\mathcal{F}_j$, we can use Lemma 2.7 of Bai and Silverstein (1998) to arrive at

$$\begin{aligned} E(K_{d,s,m}^{ij}(m-\ell+1)^4 | \mathcal{F}_{-j} \cup \mathcal{F}_{\tau_j}^{\boldsymbol{\sigma}}) &= O(\left\|\mathbf{A}(t_s^{j_m}, t_s^{j_{m+1}})\right\|^4 / (\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij})^2) \\ &= O(p_f \cdot (p-b_s-1)^{-2} n^{-2} L^{-2} / L^{-1}) \\ &= O(p_f \cdot (p-b_s-1)^{-2} n^{-2} L^{-1}), \text{ so that} \\ K_{d,s,m}^{ij}(m-\ell+1)^2 &= O_P(p_f \cdot (p-b_s-1)^{-2} n^{-2} L^{-1}), \end{aligned} \tag{7.20}$$

where $p_f = 1$ if there are no pervasive factors and $p_f = p$ if there are pervasive factors. With (7.19) and

(7.20), we can apply Theorem 2.2 of van de Geer (2002) to arrive at

$$\sum_{m=1}^{p-b_s-1} g_{d,s,m}^{ij}(m-\ell-1), \sum_{m=1}^{p-b_s-1} \frac{e_{d,s}^{ij}(m-\ell)}{(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij})^{1/2}} = O_P(p^{1/2}\cdot p_f^{1/2}\cdot(p-b_s-1)^{-1}n^{-1}L^{-1/2})$$
$$= O_P(p_f^{1/2}p^{-1/2}n^{-1}L^{-1/2}).$$

The above implies that

$$J_1, J_2 = O_P((p-b_s-1)^{1/6}\cdot p_f^{1/2}p^{-1/2}n^{-1}L^{-1/2}) = O_P(p_f^{1/2}p^{-1/3}n^{-1}L^{-1/2}).$$

We also have, as $p\to\infty$,

$$J_3 = O_P(e^{-c_{d,j,s}p^{5/6}/2}p_f^{1/2}n^{-1}L^{-1/2}).$$

The above results give

$$A_d^{ij}(s) = O_P(p_f^{1/2}p^{-1/3}n^{-1}L^{-1/2}). \tag{7.21}$$

Parallel arguments show that

$$\sum_{m=1}^{p-b_s-1} g_{v,s,m}^{ij}(m-\ell-1), \sum_{m=1}^{p-b_s-1} \frac{e_{v,s}^{ij}(m-\ell)}{(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij})^{1/2}}$$
$$= O_P((p-b_s-1)^{1/2}\cdot(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{D}_m^s\mathbf{\Sigma}(t_s^{j_m},t_s^{j_{m+1}})\mathbf{D}_m^s\mathbf{p}_{ij})^{1/2}/(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij})^{1/2})$$
$$= O_P((p-b_s-1)^{1/2}\cdot(p-b_s-1)^{-1/2}n^{-1/2}L^{-1/2}/L^{-1/2})$$
$$= O_P(n^{-1/2}),$$

where p_f cancels since $\mathbf{D}_m^s$ is only a diagonal matrix of 1 or 0, and hence if pervasive factors are affecting the numerator, it has to affect the denominator too. Parallel arguments as before show that

$$A_v^{ij}(s) = O_P(p^{1/6}n^{-1/2}). \tag{7.22}$$

Combining (7.18), (7.21), (7.22), since we at most have $p^{3/2}/n\to c>0$,

$$\frac{\mathbf{p}_{ij}^{\mathrm{T}}(\mathbf{X}_{v_s}-\mathbf{X}(s))}{(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij})^{1/2}} = O_P(p^{1/6}n^{-1/2}+p^{1/2}n^{-1}) = O_P(p^{1/6}n^{-1/2}). \tag{7.23}$$

This completes the proof of the theorem, since the above rate is free of all indices. □

Lemma 3. *Let all the assumptions in Theorem 4 hold. Then with $p/n\to c>0$ when there are no pervasive*

factors, or $p^{3/2}/n \to c > 0$ *when there are pervasive factors,*

$$\max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}} \left| \frac{I_2}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} \right| = O_P(n^{-1/6}).$$

Proof of Lemma 3. For an integer $m \geq 1$, $i = 1, \ldots, p$ and $j = 1, \ldots, L$, write

$$\frac{\mathbf{p}_{ij}^{\mathrm{T}} [\widetilde{\mathbf{X}}_v, \mathbf{E}^{\mathrm{T}}]_j^{(m)} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} = \sum_{i=1}^{3} (I_{2,i} + J_i + K_i), \text{ where, defining } e(\mathbf{J}_t) = \mathbf{J}_t - \widehat{\mathbf{J}}_t,$$

$$I_{2,1} = \frac{1}{m} \sum_{s, s-m \in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{A}(v_{s-m}, v_s) \mathbf{Z}_{d,s}^j (\boldsymbol{\epsilon}(s) - \boldsymbol{\epsilon}(s-m))^{\mathrm{T}} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}},$$

$$I_{2,2} = \frac{1}{m} \sum_{s, s-m \in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(v_{s-m}, v_s)^{1/2} \mathbf{Z}_{v,s}^j (\boldsymbol{\epsilon}(s) - \boldsymbol{\epsilon}(s-m))^{\mathrm{T}} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}},$$

$$I_{2,3} = \frac{1}{m} \sum_{s, s-m \in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}} (e(\mathbf{J}_{v_s}) - e(\mathbf{J}_{v_{s-m}})) (\boldsymbol{\epsilon}(s) - \boldsymbol{\epsilon}(s-m))^{\mathrm{T}} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}},$$

$$J_1 = \frac{1}{m} \sum_{s, s-m \in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{A}(v_{s-m}, v_s) \mathbf{Z}_{d,s}^j (\mathbf{X}(s) - \mathbf{X}_{v_s} + \mathbf{X}_{v_{s-m}} - \mathbf{X}(s-m))^{\mathrm{T}} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}},$$

$$J_2 = \frac{1}{m} \sum_{s, s-m \in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(v_{s-m}, v_s)^{1/2} \mathbf{Z}_{v,s}^j (\mathbf{X}(s) - \mathbf{X}_{v_s} + \mathbf{X}_{v_{s-m}} - \mathbf{X}(s-m))^{\mathrm{T}} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}},$$

$$J_3 = \frac{1}{m} \sum_{s, s-m \in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}} (e(\mathbf{J}_{v_s}) - e(\mathbf{J}_{v_{s-m}}) (\mathbf{X}(s) - \mathbf{X}_{v_s} + \mathbf{X}_{v_{s-m}} - \mathbf{X}(s-m))^{\mathrm{T}} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}},$$

$$K_1 = \frac{1}{m} \sum_{s, s-m \in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{A}(v_{s-m}, v_s) \mathbf{Z}_{d,s}^j (e(\mathbf{J}(s)) - e(\mathbf{J}_{v_s}) - e(\mathbf{J}(s-m)) + e(\mathbf{J}_{v_{s-m}}))^{\mathrm{T}} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}},$$

$$K_2 = \frac{1}{m} \sum_{s, s-m \in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(v_{s-m}, v_s)^{1/2} \mathbf{Z}_{v,s}^j (e(\mathbf{J}(s)) - e(\mathbf{J}_{v_s}) - e(\mathbf{J}(s-m)) + e(\mathbf{J}_{v_{s-m}}))^{\mathrm{T}} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}},$$

$$K_3 = \frac{1}{m} \sum_{s, s-m \in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}} (e(\mathbf{J}_{v_s}) - e(\mathbf{J}_{v_{s-m}})) (e(\mathbf{J}(s)) - e(\mathbf{J}_{v_s}) - e(\mathbf{J}(s-m)) + e(\mathbf{J}_{v_{s-m}}))^{\mathrm{T}} \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}}.$$

Consider $g_{d,s}^{ij} = \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{A}(v_{s-m}, v_s) \mathbf{Z}_{d,s}^j \boldsymbol{\epsilon}(s)^{\mathrm{T}} \mathbf{p}_{ij}$. Then

$$E\left(\left(\frac{1}{m} \sum_{s, s-m \in S^j(m)} g_{d,s}^{ij} \right)^2 \Big| \mathcal{F}_{-j} \right) = \frac{1}{m^2} \sum_{s, s-m \in S^j(m)} E((g_{d,s}^{ij})^2 | \mathcal{F}_{-j}) + \frac{1}{m^2} \sum_{\substack{s_k, s_k + m \in S^j(m) \\ s_1 \neq s_2}} E(g_{d,s_1}^{ij} g_{d,s_2}^{ij} | \mathcal{F}_{-j}). \tag{7.24}$$

With Assumption (D1) and (E2), we can use Lemma 2.7 of Bai and Silverstein (1998) to arrive at

$$\begin{aligned}
E((g_{d,s}^{ij})^2|\mathcal{F}_{-j}) &\leq E^{1/2}((\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-m},v_s)\mathbf{Z}_{d,s}^{j})^4|\mathcal{F}_{-j})E^{1/2}((\mathbf{p}_{ij}^{\mathrm{T}}(\mathbf{\Sigma}_{\epsilon,s}^{j})^{1/2}\mathbf{Z}_{\epsilon,s}^{j})^4|\mathcal{F}_{-j}) \\
&= O(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s-m},v_s)\mathbf{A}(v_{s-m},v_s)^{\mathrm{T}}\mathbf{p}_{ij}\cdot E^{1/2}((\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}^{j}\mathbf{p}_{ij})^2|\mathcal{F}_{-j})) \\
&= O(\left\|\mathbf{A}(v_{s-m},v_s)\right\|^2\cdot\lambda_\epsilon) = O(p_f mn^{-2}L^{-2}),
\end{aligned}$$

where $p_f = p$ if there are pervasive factors, and $p_f = 1$ otherwise. Also, by Assumption (E3), since $E(\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\epsilon}(s)|\mathcal{F}_{-j}) = 0$, by Theorem 1.4 in Rio (2013) we have that

$$\begin{aligned}
E(g_{d,s_1}^{ij}g_{d,s_2}^{ij}|\mathcal{F}_{-j}) &\leq 2O(n^{-1})E^{1/2}((\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\epsilon}(s_1))^2|\mathcal{F}_{-j}) \\
&\quad\cdot E^{1/2}((\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\epsilon}(s_2)\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s_1-m},v_{s_1})\mathbf{Z}_{d,s_1}^{j}\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s_2-m},v_{s_2})\mathbf{Z}_{d,s_2}^{j})^2|\mathcal{F}_{-j}) \\
&\leq 2O(n^{-1})E^{1/2}((\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\epsilon}(s_1))^2|\mathcal{F}_{-j})\cdot E^{1/4}((\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\epsilon}(s_2))^4|\mathcal{F}_{-j}) \\
&\quad\cdot E^{1/8}((\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s_1-m},v_{s_1})\mathbf{Z}_{d,s_1}^{j})^8|\mathcal{F}_{-j})E^{1/8}((\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{A}(v_{s_2-m},v_{s_2})\mathbf{Z}_{d,s_2}^{j})^8|\mathcal{F}_{-j}) \\
&= O(n^{-1}\left\|\mathbf{A}(v_{s-m},v_s)\right\|^2) = O(p_f mn^{-3}L^{-2}),
\end{aligned}$$

where the third inequality sign used Lemma 2.7 of Bai and Silverstein (1998), and the existence of the eighth moments after applying the Cauchy-Schwarz inequality. Using these two results, (7.24) becomes

$$E\bigg(\Big(\frac{1}{m}\sum_{s,s-m\in S^j(m)}g_{d,s}^{ij}\Big)^2\Big|\mathcal{F}_{-j}\bigg) = O(m^{-2}p_f mn^{-1}L^{-2}) = O(p_f m^{-1}n^{-1}L^{-2}).$$

This implies that

$$I_{2,1} = O_P(p_f^{1/2}m^{-1/2}n^{-1/2}L^{-1}/L^{-1}) = O_P(p_f^{1/2}m^{-1/2}n^{-1/2}). \tag{7.25}$$

Now consider $g_{v,s}^{ij} = \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(v_{s-m},v_s)^{1/2}\mathbf{Z}_{v,s}^{j}\boldsymbol{\epsilon}(s)^{\mathrm{T}}\mathbf{p}_{ij}/\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}$. Parallel arguments using Assumption (V1) and (E2) give

$$\begin{aligned}
E((g_{v,s}^{ij})^2|\mathcal{F}_{-j}\cup\mathcal{F}_{\tau_j}^{\boldsymbol{\sigma}}) &= O(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(v_{s-m},v_s)\mathbf{p}_{ij}/(\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij})^2) = O(p_f mn^{-1}L^{-1}/(p_f^2L^{-2})) \\
&= O(p_f^{-1}mn^{-1}L), \\
E(g_{v,s_1}^{ij}g_{v,s_2}^{ij}|\mathcal{F}_{-j}\cup\mathcal{F}_{\tau_j}^{\boldsymbol{\sigma}}) &= O(n^{-1}p_f^{-1}mn^{-1}L) = O(p_f^{-1}mn^{-2}L)).
\end{aligned}$$

Hence using decomposition parallel to (7.24),

$$I_{2,2} = O_P(m^{-2}\cdot p_f^{-1}mL)^{1/2} = O_P(p_f^{-1/2}m^{-1/2}L^{1/2}). \tag{7.26}$$

For terms involving jumps, using Assumption (W1) to (W3) and the rate in Fan and Wang (2007), we have

$$\begin{aligned} I_{2,3} &= O_P(n^{-1/4}L^{3/4}), \\ J_3 &= O_P(n^{-1/4}L^{1/4}\cdot p^{1/6}n^{-1/2}) = O_P(p^{1/6}n^{-3/4}L^{1/4}), \\ K_1 &= O_P(\|\mathbf{A}(v_{s-m},v_s)\|/L^{-1}\cdot n^{-1/4}L^{-1/4}) = O_P(p_f^{1/2}m^{1/2}n^{-5/4}L^{-1/4}), \\ K_2 &= O_P(p_f^{-1/2}m^{1/2}n^{-1/2}L^{1/2}\cdot n^{-1/4}L^{-1/4}) = O_P(p_f^{-1/2}m^{1/2}n^{-3/4}L^{1/4}), \\ K_3 &= O_P(n^{-1/2}L^{1/2}), \end{aligned} \tag{7.27}$$

where J_3 used the result of Lemma 2. Using the result of Lemma 2 again, we have

$$\begin{aligned} J_1 &= O_P(nm^{-1}\cdot p_f^{1/2}m^{1/2}n^{-1}L^{-1/2}\cdot p^{1/6}n^{-1/2}) = O_P(p_f^{1/2}p^{1/6}m^{-1/2}n^{-1/2}L^{-1/2}), \\ J_2 &= O_P(nm^{-1}\cdot m^{1/2}n^{-1/2}\cdot p^{1/6}n^{-1/2}) = O_P(m^{-1/2}p^{1/6}). \end{aligned} \tag{7.28}$$

At $m = K \asymp n^{2/3}$, (7.25), (7.26), (7.27) and (7.28) imply that, for $p_f = 1$ with $p \asymp n$ or $p_f = p \asymp n^{2/3}$,

$$\frac{\mathbf{p}_{ij}^{\mathrm{T}}[\widetilde{\mathbf{X}}_v, \mathbf{E}^{\mathrm{T}}]_j^{(K)}\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}} = O_P(n^{-1/6}).$$

At $m = 1$, (7.25), (7.26), (7.27) and (7.28) imply that, for $p_f = 1$ with $p \asymp n$ or $p_f = p \asymp n^{2/3}$,

$$\frac{\mathbf{p}_{ij}^{\mathrm{T}}[\widetilde{\mathbf{X}}_v, \mathbf{E}^{\mathrm{T}}]_j^{(1)}\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}} = O_P(p^{1/6}).$$

Since the above two results are free of all indices, they imply that

$$\max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}}\left|\frac{I_2}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\right| = O_P(n^{-1/6}+p^{1/6}K^{-1}) = O_P(n^{-1/6}).$$

This completes the proof of the lemma. □

Lemma 4. *Let all the assumptions in Theorem 4 hold. Then with $p/n \to c > 0$ when there are no pervasive factors, or $p^{3/2}/n \to c > 0$ when there are pervasive factors,*

$$\max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}}\left|\frac{I_3}{\mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}\right| = O_P(n^{-1/6}).$$

Proof of Lemma 4. Consider for an integer $m \geq 1$ and $i = 1,\ldots,p$, $j = 1,\ldots,L$, using the notations

in the proof of Lemma 3,

$$\frac{\mathbf{p}_{ij}^{\mathrm{T}}[\mathbf{E},\mathbf{E}^{\mathrm{T}}]_j^{(m)}\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}} = \sum_{\ell=1}^{3} I_{3,\ell} + 2\sum_{\ell=1}^{3} I_{3,\ell}, \text{ where}$$

$$I_{3,1}(m) = \frac{1}{m}\sum_{s,s-m\in S^j(m)} \frac{(\mathbf{p}_{ij}^{\mathrm{T}}(\boldsymbol{\epsilon}(s)-\boldsymbol{\epsilon}(s-m)))^2}{\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}},$$

$$I_{3,2} = \frac{1}{m}\sum_{s,s-m\in S^j(m)} \frac{(\mathbf{p}_{ij}^{\mathrm{T}}(\mathbf{X}(s)-\mathbf{X}_{v_s}+\mathbf{X}_{v_{s-m}}-\mathbf{X}(s-m)))^2}{\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}},$$

$$I_{3,3} = \frac{1}{m}\sum_{s,s-m\in S^j(m)} \frac{(\mathbf{p}_{ij}^{\mathrm{T}}(e(\mathbf{J}(s))-e(\mathbf{J}_{v_s})-e(\mathbf{J}(s-m))+e(\mathbf{J}_{v_{s-m}})))^2}{\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}},$$

$$I_{3,4} = \frac{1}{m}\sum_{s,s-m\in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}}(\boldsymbol{\epsilon}(s)-\boldsymbol{\epsilon}(s-m))(\mathbf{X}(s)-\mathbf{X}_{v_s}+\mathbf{X}_{v_{s-m}}-\mathbf{X}(s-m))^{\mathrm{T}}\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}},$$

$$I_{3,5} = \frac{1}{m}\sum_{s,s-m\in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}}(\boldsymbol{\epsilon}(s)-\boldsymbol{\epsilon}(s-m))(e(\mathbf{J}(s))-e(\mathbf{J}_{v_s})-e(\mathbf{J}(s-m))+e(\mathbf{J}_{v_{s-m}}))^{\mathrm{T}}\mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}},$$

$$I_{3,6} = \frac{1}{m}\sum_{s,s-m\in S^j(m)} \frac{\mathbf{p}_{ij}^{\mathrm{T}}(\mathbf{X}(s)-\mathbf{X}_{v_s}+\mathbf{X}_{v_{s-m}}-\mathbf{X}(s-m))}{\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}}$$
$$\cdot (e(\mathbf{J}(s))-e(\mathbf{J}_{v_s})-e(\mathbf{J}(s-m))+e(\mathbf{J}_{v_{s-m}}))^{\mathrm{T}}\mathbf{p}_{ij}.$$

We consider $I_{3,2}$ first, which by Lemma 2 has

$$I_{3,2} = O_P(nm^{-1}\cdot p^{1/3}n^{-1}) = O_P(p^{1/3}m^{-1}).$$

Using Assumption (W1) to (W3) and the rate of wavelet removal in Fan and Wang (2007), we have

$$I_{3,3} = O_P(n^{-1/2}L^{1/2}),$$
$$I_{3,5} = O_P(n^{-1/4}L^{3/4}),$$
$$I_{3,6} = O_P(p^{1/6}n^{-1/2}\cdot n^{-1/4}L^{1/4}) = O_P(p^{1/6}n^{-3/4}L^{1/4}).$$

Consider $h_s^{ij} = \mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\epsilon}(s)(\mathbf{X}(s)-\mathbf{X}_{v_s})^{\mathrm{T}}\mathbf{p}_{ij}/\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\mathbf{p}_{ij}$. Then using Assumption (E3), (D1) and (V1) that eighth moments exist, with $s_1 \neq s_2$,

$$E((h_s^{ij})^2|\mathcal{F}_{-j}\cup\mathcal{F}_{\tau_j}^{\boldsymbol{\sigma}}) = O(p^{1/3}n^{-1}L),$$
$$E(h_{s_1}^{ij}h_{s_2}^{ij}|\mathcal{F}_{-j}\cup\mathcal{F}_{\tau_j}^{\boldsymbol{\sigma}}) = O(n^{-1}\cdot L\cdot p^{1/3}n^{-1}) = O(p^{1/3}n^{-2}L).$$

Hence using decomposition parallel to (7.24), we can conclude that

$$I_{3,4} = O_P(m^{-2} \cdot n \cdot p^{1/3} n^{-1} L + m^{-2} \cdot n^2 \cdot p^{1/3} n^{-2} L)^{1/2} = O_P(p^{1/6} m^{-1} L^{1/2}).$$

Finally, for $K \asymp n^{2/3}$, we consider the rate of

$$\begin{aligned}
(\mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}) \left(I_{3,1}(K) - \frac{|S^j(K)|_K}{|S^j(1)|} I_{3,1}(1) \right) &= J_1 - 2J_2 + J_3, \text{ where} \\
J_1 = \frac{1}{K} \sum_{s, s-K \in S^j(K)} (\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\epsilon}(s))^2 - \frac{|S^j(K)|_K}{|S^j(1)|} \sum_{s, s-1 \in S^j(1)} (\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\epsilon}(s))^2, \\
J_2 = \frac{1}{K} \sum_{s, s-K \in S^j(K)} \mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\epsilon}(s) \boldsymbol{\epsilon}(s-K)^{\mathrm{T}} \mathbf{p}_{ij} - \frac{|S^j(K)|_K}{|S^j(1)|} \sum_{s, s-1 \in S^j(1)} \mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\epsilon}(s) \boldsymbol{\epsilon}(s-1)^{\mathrm{T}} \mathbf{p}_{ij}, \\
J_3 = \frac{1}{K} \sum_{s, s-K \in S^j(K)} (\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\epsilon}(s-K))^2 - \frac{|S^j(K)|_K}{|S^j(1)|} \sum_{s, s-1 \in S^j(1)} (\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\epsilon}(s-1))^2.
\end{aligned}$$

With Assumption (E1) to (E3), writing $g_{m,s}^{ij} = \mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\epsilon}(s) \boldsymbol{\epsilon}(s-m)^{\mathrm{T}} \mathbf{p}_{ij}$, by Lemma 2.7 of Bai and Silverstein (1998),

$$\begin{aligned}
E\left\{ \Big(\frac{1}{m} \sum_{s, s-m \in S^j(m)} g_{m,s}^{ij} \Big)^2 \Big| \mathcal{F}_{-j} \right\} = O(m^{-2} n \cdot 1 + n^{-1} \cdot m^{-2} n^2 \cdot 1) = O(m^{-2} n), \text{ hence} \\
\frac{1}{m} \sum_{s, s-m \in S^j(m)} g_{m,s}^{ij} = O_P(m^{-1} n^{1/2}),
\end{aligned}$$

which implies that

$$J_2 = O_P(K^{-1} n^{1/2}) = O_P(n^{-1/6}).$$

We can further decompose $J_1 = J_{11} - J_{12} + J_{13}$, where

$$\begin{aligned}
J_{11} &= \frac{1}{K} \sum_{s, s-K \in S^j(K)} ((\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\epsilon}(s))^2 - \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}_{\epsilon,s}^j \mathbf{p}_{ij}), \\
J_{12} &= \frac{|S^j(K)|_K}{|S^j(1)|} \sum_{s, s-1 \in S^j(1)} ((\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\epsilon}(s))^2 - \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}_{\epsilon,s}^j \mathbf{p}_{ij}), \\
J_{13} &= \frac{1}{K} \sum_{s, s-K \in S^j(K)} \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}_{\epsilon,s}^j \mathbf{p}_{ij} - \frac{|S^j(K)|_K}{|S^j(1)|} \sum_{s, s-1 \in S^j(1)} \mathbf{p}_{ij}^{\mathrm{T}} \mathbf{\Sigma}_{\epsilon,s}^j \mathbf{p}_{ij}.
\end{aligned}$$

Consider

$$
\begin{aligned}
J_{13} &= \frac{1}{K}\sum_{s,s-K\in S^j(K)} \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}^j\mathbf{p}_{ij} - \frac{1}{K}\sum_{s,s-1\in S^j(1)} \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}^j\mathbf{p}_{ij} + \frac{K-1}{Kn(j)}\sum_{s,s-1\in S^j(1)} \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}\mathbf{p}_{ij} \\
&= -\frac{1}{K}\sum_{s=1}^{K-1} \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}^j\mathbf{p}_{ij} + \frac{K-1}{Kn(j)}\sum_{s=1}^{n(j)} \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}^j\mathbf{p}_{ij} \\
&= \left(\mathbf{p}_{ij}^{\mathrm{T}}E(\mathbf{\Sigma}_{\epsilon,s}^j)\mathbf{p}_{ij} - \frac{1}{K}\sum_{s=1}^{K-1} \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}^j\mathbf{p}_{ij}\right) + \left(\frac{1}{n(j)}\sum_{s=1}^{n(j)} \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}^j\mathbf{p}_{ij} - \mathbf{p}_{ij}^{\mathrm{T}}E(\mathbf{\Sigma}_{\epsilon,s}^j)\mathbf{p}_{ij}\right) \\
&- \frac{1}{Kn(j)}\sum_{s=1}^{n(j)} \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}^j\mathbf{p}_{ij} \\
&= O_P(K^{-1/2}) + O_P(n^{-1/2}) + O_P(K^{-1}) = O_P(n^{-1/3}),
\end{aligned}
$$

where the last line used the weak law of large number given $\mathcal{F}_{-j}$.

Now define $g_s^{ij} = (\mathbf{p}_{ij}^{\mathrm{T}}\boldsymbol{\epsilon}(s))^2 - \mathbf{p}_{ij}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon,s}^j\mathbf{p}_{ij}$. Using Lemma 2.7 of Bai and Silverstein (1998) under Assumption (E1) to (E3), we have

$$
\begin{aligned}
E(J_{11}^2|\mathcal{F}_{-j}\cup\{\mathbf{\Sigma}_{\epsilon,u}, u\in[0,1]\}) &= K^{-2}\sum_{s,s-K\in S^j(K)} E((g_s^{ij})^2|\mathcal{F}_{-j}\cup\{\mathbf{\Sigma}_{\epsilon,u}, u\in[0,1]\}) \\
&+ K^{-2}\sum_{s_1\neq s_2} E(g_{s_1}^{ij}g_{s_2}^{ij}|\mathcal{F}_{-j}\cup\{\mathbf{\Sigma}_{\epsilon,u}, u\in[0,1]\}) \\
&= O(K^{-2}n\cdot 1 + K^{-2}n^2\cdot n^{-1}\cdot 1) = O(n^{-1/3}).
\end{aligned}
$$

The above implies that

$$J_{11} = O_P(n^{-1/6}) = J_{12}.$$

The rates for J_{11}, J_{12} and J_{13} imply that

$$J_1 = O_P(n^{-1/6}) = J_3,$$

so that combining with the rate of J_2, we have

$$I_{3,1}(K) - \frac{|S^j(K)|_K}{|S^j(1)|}I_{3,1}(1) = O_P(n^{-1/6}L).$$

Finally, among $I_{3,2}$ to $I_{3,6}$, when $m = K \asymp n^{2/3}$, the dominating term is $I_{3,5} = O_P(n^{-1/4}L^{3/4})$, while it is

$I_{3,2} = O_P(p^{1/3})$ when $m = 1$. Hence

$$\max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}} \left| \frac{I_3}{\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} \right| = O_P(n^{-1/6} L) + O_P(n^{-1/4} L^{3/4}) + O_P(K^{-1} \cdot p^{1/3})$$
$$= O_P(n^{-1/6} L) = O_P(n^{-1/6}),$$

since L is finite. This completes the proof of the lemma. □

Proof of Theorem 2, 4. Combining the results of Lemma 1, 3 and 4, we have

$$\max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}} \left| \frac{\mathbf{p}_{ij}^{\mathrm{T}} \widetilde{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}}{\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} - 1 \right| \leq \max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}} \left| \frac{I_1}{\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} - 1 \right| + 2 \max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}} \left| \frac{I_2}{\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} \right|$$
$$+ \max_{\substack{i=1,\ldots,p\\ j=1,\ldots,L}} \left| \frac{I_3}{\mathbf{p}_{ij}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{p}_{ij}} \right| = O_P(n^{-1/6}).$$

Note that the above result is equivalent to the first main result in Theorem 2. For the second main result,

$$\left\| \widehat{\boldsymbol{\Sigma}}(0,1) \boldsymbol{\Sigma}_{\mathrm{Ideal}}(0,1)^{-1} - \mathbf{I}_p \right\| = \left\| \sum_{j=1}^{L} (\widehat{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j) \boldsymbol{\Sigma}_{\mathrm{Ideal}}(\tau_{j-1}, \tau_j)^{-1} - \mathbf{I}_p) \boldsymbol{\Sigma}_{\mathrm{Ideal}}(\tau_{j-1}, \tau_j) \boldsymbol{\Sigma}_{\mathrm{Ideal}}(0,1)^{-1} \right\|$$
$$\leq \sum_{j=1}^{L} \left\| \widehat{\boldsymbol{\Sigma}}(\tau_{j-1}, \tau_j) \boldsymbol{\Sigma}_{\mathrm{Ideal}}(\tau_{j-1}, \tau_j)^{-1} - \mathbf{I}_p \right\|$$
$$\cdot \left\| \mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{P}_{-j}) \cdot \left(\mathrm{diag}(\mathbf{P}_{-j}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{P}_{-j}) + \sum_{i \neq j} \mathbf{P}_{-j}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathrm{Ideal}}(\tau_{i-1}, \tau_i) \mathbf{P}_{-j} \right)^{-1} \right\|$$
$$= O_P\left(L n^{-1/6} \cdot \max_{j=1,\ldots,L} \left\| \left(\mathbf{I}_p + \sum_{i \neq j} \mathbf{P}_{-j}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathrm{Ideal}}(\tau_{i-1}, \tau_i) \mathbf{P}_{-j} \mathrm{diag}^{-1}(\mathbf{P}_{-j}^{\mathrm{T}} \boldsymbol{\Sigma}(\tau_{j-1}, \tau_j) \mathbf{P}_{-j}) \right)^{-1} \right\| \right)$$
$$= O_P(n^{-1/6}).$$

Hence these complete the proof of Theorem 2 and the equivalent parts in Theorem 4 under jumps removed data.

To complete the proof of Theorem 4, note that for a generic constant $C > 0$,

$$\left\| \sum_{0 \leq t \leq 1} (\Delta \mathbf{J}_t \Delta \mathbf{J}_t^{\mathrm{T}} - \Delta \widehat{\mathbf{J}}_t \Delta \widehat{\mathbf{J}}_t^{\mathrm{T}}) \right\| \leq C \max_{0 \leq t \leq 1} \left\| \Delta \mathbf{J}_t \Delta \mathbf{J}_t^{\mathrm{T}} - \Delta \widehat{\mathbf{J}}_t \Delta \widehat{\mathbf{J}}_t^{\mathrm{T}}) \right\|$$
$$\leq 2C \max_{0 \leq t \leq 1} \left\| \Delta \mathbf{J}_t - \Delta \widehat{\mathbf{J}}_t \right\| \cdot \left\| \Delta \mathbf{J}_t \right\| + C \max_{0 \leq t \leq 1} \left\| \Delta \mathbf{J}_t - \Delta \widehat{\mathbf{J}}_t \right\|^2$$
$$= O_P(n^{-1/4} L^{-1/4}),$$

where the first line used Assumption (W2) that there are only finite number of jumps in $[0, 1]$ for each stock, and the second line used Assumption (W3) that there are only finite number of cojumps, with rate

of jumps removal given as in Fan and Wang (2007). This completes the proof of Theorem 4. □

Proof of Theorem 5. Define $\mathbf{D}_j = \text{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\mathbf{P}_{-j})$ and $\widetilde{\mathbf{D}}_j = \text{diag}(\mathbf{P}_{-j}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}(\tau_{j-1},\tau_j)\mathbf{P}_{-j})$. Define $\mathbf{e}_i$ to be the unit vector with 1 on the ith position and 0 elsewhere, and $\big\|\mathbf{A}\big\|_1 = \max_j \sum_i |a_{ij}|$ the L_1 norm of a matrix $\mathbf{A}$. Then for some $i = 1, \ldots, p$,

$$
\begin{aligned}
p^{1/2}\big\|\widehat{\mathbf{w}}_{\text{opt}}\big\|_{\max} &= \frac{p^{1/2}|\mathbf{e}_i^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}\mathbf{1}_p|}{\mathbf{1}_p^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}\mathbf{1}_p} \leq \frac{p^{1/2}\big\|\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}\big\|_1}{p\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}(0,1)^{-1})} \leq \frac{p^{1/2}\cdot p^{1/2}/\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}(0,1))}{p/\lambda_{\max}(\widehat{\boldsymbol{\Sigma}}(0,1))} \\
&\leq \frac{\sum_{j=1}^L \lambda_{\max}(\widetilde{\mathbf{D}}_j)}{\sum_{j=1}^L \lambda_{\min}(\widetilde{\mathbf{D}}_j)} \\
&\leq \frac{L\max_{1\leq j\leq L}\lambda_{\max}(\widetilde{\mathbf{D}}_j\mathbf{D}_j^{-1}-\mathbf{I}_p)\lambda_{\max}(\mathbf{D}_j)+\sum_{j=1}^L\lambda_{\max}(\mathbf{D}_j)}{L\min_{1\leq j\leq L}\lambda_{\min}(\widetilde{\mathbf{D}}_j\mathbf{D}_j^{-1}-\mathbf{I}_p)\lambda_{\min}(\mathbf{D}_j)+\sum_{j=1}^L\lambda_{\min}(\mathbf{D}_j)} \\
&\leq \frac{(\max_{1\leq j\leq L}\lambda_{\max}(\widetilde{\mathbf{D}}_j\mathbf{D}_j^{-1}-\mathbf{I}_p)+1)\max_{1\leq j\leq L}\lambda_{\max}(\mathbf{D}_j)}{(\min_{1\leq j\leq L}\lambda_{\min}(\widetilde{\mathbf{D}}_j\mathbf{D}_j^{-1}-\mathbf{I}_p)+1)\min_{1\leq j\leq L}\lambda_{\min}(\mathbf{D}_j)} \\
&\xrightarrow{\text{P}} \frac{\max_{1\leq j\leq L}\lambda_{\max}(\mathbf{D}_j)}{\min_{1\leq j\leq L}\lambda_{\min}(\mathbf{D}_j)} \leq \frac{\max_{1\leq j\leq L}\lambda_{\max}(\boldsymbol{\Sigma}(\tau_{j-1},\tau_j))}{\min_{1\leq j\leq L}\lambda_{\min}(\boldsymbol{\Sigma}(\tau_{j-1},\tau_j))},
\end{aligned}
$$

where the last line follows from the results of Theorem 2 and Theorem 4. For the theoretical minimum-variance portfolio,

$$
\begin{aligned}
p^{1/2}\big\|\mathbf{w}_{\text{theo}}\big\|_{\max} &= \frac{p^{1/2}|\mathbf{e}_i^{\mathrm{T}}\boldsymbol{\Sigma}(0,1)^{-1}\mathbf{1}_p|}{\mathbf{1}_p^{\mathrm{T}}\boldsymbol{\Sigma}(0,1)^{-1}\mathbf{1}_p} \leq \frac{p^{1/2}\big\|\boldsymbol{\Sigma}(0,1)^{-1}\big\|_1}{p\lambda_{\min}(\boldsymbol{\Sigma}(0,1)^{-1})} \leq \frac{p^{1/2}\cdot p^{1/2}/\lambda_{\min}(\boldsymbol{\Sigma}(0,1))}{p/\lambda_{\max}(\boldsymbol{\Sigma}(0,1))} \\
&\leq \frac{\sum_{j=1}^L \lambda_{\max}(\mathbf{D}_j)}{\sum_{j=1}^L \lambda_{\min}(\mathbf{D}_j)} = \frac{\max_{1\leq j\leq L}\lambda_{\max}(\mathbf{D}_j)}{\min_{1\leq j\leq L}\lambda_{\min}(\mathbf{D}_j)} \leq \frac{\max_{1\leq j\leq L}\lambda_{\max}(\boldsymbol{\Sigma}(\tau_{j-1},\tau_j))}{\min_{1\leq j\leq L}\lambda_{\min}(\boldsymbol{\Sigma}(\tau_{j-1},\tau_j))}.
\end{aligned}
$$

For the actual risk bound, define $\mathbf{R} = \sum_{j=1}^L \mathbf{P}_{-j}(\widetilde{\mathbf{D}}_j\mathbf{D}_j^{-1} - \mathbf{I}_p)\mathbf{D}_j\mathbf{P}_{-j}^{\mathrm{T}}$. We first consider the case of no pervasive factors. Consider

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}(0,1)^{-1} &= \Big(\sum_{j=1}^L \mathbf{P}_{-j}\widetilde{\mathbf{D}}_j\mathbf{P}_{-j}^{\mathrm{T}}\Big)^{-1} = \Big(\sum_{j=1}^L \mathbf{P}_{-j}(\widetilde{\mathbf{D}}_j\mathbf{D}_j^{-1}-\mathbf{I}_p)\mathbf{D}_j\mathbf{P}_{-j}^{\mathrm{T}} + \sum_{j=1}^L \mathbf{P}_{-j}\mathbf{D}_j\mathbf{P}_{-j}^{\mathrm{T}}\Big)^{-1} \\
&= (\mathbf{I}_p + \boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\mathbf{R})^{-1}\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1} \\
&= \boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1} + \sum_{k\geq 1}\big(-\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\mathbf{R}\big)^k\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1},
\end{aligned}
$$

where the Neumann's series expansion in the last line is valid since

$$\begin{aligned}\sum_{k\geq 0}\left\|\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\right\|^k\|\mathbf{R}\|^k &\leq 1+\sum_{k\geq 1}\frac{\|\mathbf{R}\|^k}{\lambda_{\min}^k(\boldsymbol{\Sigma}_{\text{Ideal}}(0,1))}\\ &\leq 1+\sum_{k\geq 1}\frac{L^k\max_{1\leq j\leq L}\left\|\widetilde{\mathbf{D}}_j\mathbf{D}_j^{-1}-\mathbf{I}_p\right\|^k\max_{1\leq j\leq L}\left\|\boldsymbol{\Sigma}(\tau_{j-1},\tau_j)\right\|^k}{L^k\min_{1\leq j\leq L}\lambda_{\min}^k(\boldsymbol{\Sigma}(\tau_{j-1},\tau_j))}\\ &\xrightarrow{\text{P}} 1<\infty,\end{aligned}$$

where the last line follows from the results in Theorem 2 and 4. This implies that, in probability,

$$\left\|\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}-\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\right\|\leq\lambda_{\max}(\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1})\sum_{k\geq 1}\frac{\|\mathbf{R}\|^k}{\lambda_{\min}^k(\boldsymbol{\Sigma}_{\text{Ideal}}(0,1))}\xrightarrow{\text{P}}0. \tag{7.29}$$

With the above, consider the decomposition $pR(\widehat{\mathbf{w}}_{\text{opt}})=I_1+I_2+I_3$, where

$$\begin{aligned}I_1&=\frac{p\mathbf{1}_p^{\text{T}}(\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}-\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1})\boldsymbol{\Sigma}(0,1)\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}\mathbf{1}_p}{(\mathbf{1}_p^{\text{T}}\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}\mathbf{1}_p)^2},\\ I_2&=\frac{p\mathbf{1}_p^{\text{T}}\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\boldsymbol{\Sigma}(0,1)(\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}-\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1})\mathbf{1}_p}{(\mathbf{1}_p^{\text{T}}\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}\mathbf{1}_p)^2},\\ I_3&=\frac{p\mathbf{1}_p^{\text{T}}\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\boldsymbol{\Sigma}(0,1)\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\mathbf{1}_p}{(\mathbf{1}_p^{\text{T}}\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}\mathbf{1}_p)^2}.\end{aligned}$$

By (7.29), with $\left\|\boldsymbol{\Sigma}(0,1)\right\|\leq C$ where C is a generic constant since there are no pervasive factors,

$$|I_1|\leq\frac{p^2\left\|\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}-\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\right\|\cdot C\cdot(\left\|\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}-\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\right\|+\lambda_{\max}(\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}))}{p^2\big(\lambda_{\min}(\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1})-\left\|\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}-\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\right\|\big)^2}\xrightarrow{\text{P}}0.$$

Similarly, $|I_2|\xrightarrow{\text{P}}0$. For I_3, by (7.29),

$$\begin{aligned}|I_3|&\leq\frac{p^2\lambda_{\max}^2(\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1})\lambda_{\max}(\boldsymbol{\Sigma}(0,1))}{p^2\big(\lambda_{\min}(\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1})-\left\|\widehat{\boldsymbol{\Sigma}}(0,1)^{-1}-\boldsymbol{\Sigma}_{\text{Ideal}}(0,1)^{-1}\right\|\big)^2}\\ &\xrightarrow{\text{P}}\frac{\lambda_{\max}^2(\boldsymbol{\Sigma}_{\text{Ideal}}(0,1))}{\lambda_{\min}^2(\boldsymbol{\Sigma}_{\text{Ideal}}(0,1))}\lambda_{\max}(\boldsymbol{\Sigma}(0,1))\\ &\leq\left(\frac{\sum_{j=1}^L\lambda_{\max}(\boldsymbol{\Sigma}(\tau_{j-1},\tau_j))}{\sum_{j=1}^L\lambda_{\min}(\boldsymbol{\Sigma}(\tau_{j-1},\tau_j))}\right)^2\lambda_{\max}(\boldsymbol{\Sigma}(0,1))\\ &=\left(\frac{\max_{1\leq j\leq L}\lambda_{\max}(\boldsymbol{\Sigma}(\tau_{j-1},\tau_j))}{\min_{1\leq j\leq L}\lambda_{\min}(\boldsymbol{\Sigma}(\tau_{j-1},\tau_j))}\right)^2\lambda_{\max}(\boldsymbol{\Sigma}(0,1)),\end{aligned}$$

which leads to the result in the theorem.

If there are pervasive factors, abbreviating $\mathbf{\Sigma}(0,1)$ as $\mathbf{\Sigma}$ etc, consider

$$\begin{aligned} R(\widehat{\mathbf{w}}_{\text{opt}}) &= \frac{\mathbf{1}_p^{\text{T}}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{1}_p}{(\mathbf{1}_p^{\text{T}}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{1}_p)^2} \leq \frac{\lambda_{\max}(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma})}{\mathbf{1}_p^{\text{T}}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{1}_p} \leq \frac{\lambda_{\max}(\widehat{\mathbf{\Sigma}})\lambda_{\max}(\mathbf{\Sigma})}{p\lambda_{\min}(\widehat{\mathbf{\Sigma}})} \\ &= O_P(\lambda_{\max}(\mathbf{\Sigma})), \end{aligned}$$

where the last line follows from the results in Theorem 2 and 4. For the actual risk bound for $\mathbf{w}_{\text{theo}}$,

$$pR(\mathbf{w}_{\text{theo}}) = \frac{p}{\mathbf{1}_p^{\text{T}}\mathbf{\Sigma}(0,1)^{-1}\mathbf{1}_p} \leq \lambda_{\max}(\mathbf{\Sigma}(0,1)).$$

This completes the proof of the theorem. □

References

Abadir, K. M., Distaso, W., and Žikeš, F. (2014). Design-free estimation of variance matrices. *Journal of Econometrics*, 181(2):165 – 180.

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., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18(2):351–416.

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.

Asparouhova, E., Bessembinder, H., and Kalcheva, I. (2013). Noisy prices and inference regarding returns. *The Journal of Finance*, 68(2):665–714.

Bai, Z. and Silverstein, J. (2009). *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics. Springer New York.

Bai, Z. and Silverstein, J. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics, New York, 2 edition.

Bai, Z. D., Miao, B. Q., and Pan, G. M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *The Annals of Probability*, 35(4):1532–1572.

Bai, Z. D. and 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.

Barndorff-Nielsen, O., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal*, 12(3):C1–C32.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011a). 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., Hansen, P. R., Lunde, A., and Shephard, N. (2011b). 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.

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

Chen, R. Y. and Mykland, P. A. (2017). Model-free approaches to discern non-stationary microstructure noise and time-varying liquidity in high-frequency data. *Journal of Econometrics*, 200(1):79 – 103.

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.

Dao, C., Lu, K., and Xiu, D. (2017). Knowing factors or factor loadings, or neither? evaluating estimators of large covariance matrices with noisy and asynchronous data. *Chicago Booth Research Paper No. 17-02*.

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.

Fan, J. and Kim, D. (2017). Robust high-dimensional volatility matrix estimation for high-frequency factor model. *Journal of the American Statistical Association*. Forthcoming.

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. 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.

Gilder, D., Shackleton, M. B., and Taylor, S. J. (2014). Cojumps in stock prices: Empirical evidence. *Journal of Banking and Finance*, 40(Supplement C):443 – 459.

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

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 361–379, Berkeley, Calif. University of California Press.

Kim, D., Wang, Y., and Zou, J. (2016). Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stochastic Processes and their Applications*, 126(11):3527 – 3577.

Lam, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *Ann. Statist.*, 44(3):928–953.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060.

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

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457.

Rio, E. (2013). Inequalities and limit theorems for weakly dependent sequences. Lecture.

Tao, M., Wang, Y., and Chen, X. (2013). Fast convergence rates in estimating large volatility matrices using high-frequency financial data. *Econometric Theory*, 29(4):838856.

Tao, M., Wang, Y., Yao, Q., and Zou, J. (2011). Large volatility matrix inference via combining low-frequency and high-frequency approaches. *Journal of the American Statistical Association*, 106(495):1025–1040.

van de Geer, S. A. (2002). On hoeffding's inequality for dependent random variables. In Dehling, H., Mikosch, T., and Sørensen, M., editors, *Empirical Process Techniques for Dependent Data*, pages 161–169, Boston, MA. Birkhäuser Boston.

Wang, Y. and Zou, J. (2010). Vast volatility matrix estimation for high-frequency financial data. *Ann. Statist.*, 38(2):943–978.

Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*, 159(1):235 – 250.

Xue, Y., Gençay, R., and Fagan, S. (2014). Jump detection with wavelets for high-frequency financial time series. *Quantitative Finance*, 14(8):1427–1444.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33 – 47.