Journal of the American Statistical Association

Vast Volatility Matrix Estimation Using High-Frequency Data for Portfolio Selection

Jianqing Fan [a] , Yingying Li [b] & Ke Yu [a]

[a] Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, 08544, USA

[b] Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

PLEASE SCROLL DOWN FOR ARTICLE

Vast Volatility Matrix Estimation Using High-Frequency Data for Portfolio Selection

Jianqing FAN, Yingying LI, and Ke YU

Portfolio allocation with gross-exposure constraint is an effective method to increase the efficiency and stability of portfolios selection among a vast pool of assets, as demonstrated by Fan, Zhang, and Yu. The required high-dimensional volatility matrix can be estimated by using high-frequency financial data. This enables us to better adapt to the local volatilities and local correlations among a vast number of assets and to increase significantly the sample size for estimating the volatility matrix. This article studies the volatility matrix estimation using high-dimensional, high-frequency data from the perspective of portfolio selection. Specifically, we propose the use of "pairwise-refresh time" and "all-refresh time" methods based on the concept of "refresh time" proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard for the estimation of vast covariance matrix and compare their merits in the portfolio selection. We establish the concentration inequalities of the estimates, which guarantee desirable properties of the estimated volatility matrix in vast asset allocation with gross-exposure constraints. Extensive numerical studies are made via carefully designed simulations. Comparing with the methods based on low-frequency daily data, our methods can capture the most recent trend of the time varying volatility and correlation, hence provide more accurate guidance for the portfolio allocation in the next time period. The advantage of using high-frequency data is significant in our simulation and empirical studies, which consist of 50 simulated assets and 30 constituent stocks of Dow Jones Industrial Average index.

KEY WORDS: Concentration inequalities; High-frequency data; Portfolio allocation; Refresh time; Risk assessment; Volatility matrix estimation.

1. INTRODUCTION

The mean-variance efficient portfolio theory by Markowitz (1952, 1959) has had a profound impact on modern finance. Yet, its applications to practical portfolio selection face a number of challenges. It is well known that the selected portfolios depend too sensitively on the expected future returns and volatility matrix. This leads to the puzzle postulated by Jagannathan and Ma (2003) why no short-sale portfolio outperforms the efficient Markowicz portfolio. The sensitivity on the dependence can be effectively addressed by the introduction of the constraint on the gross exposure of portfolios (Fan, Zhang, and Yu 2011). Their results not only give a theoretical answer to the puzzle postulated by Jagannathan and Ma (2003) but also pave a way for optimal portfolio selection in practice.

The second challenge of the implementation of Markowitz's portfolio selection theory is the intrinsic difficulty of the estimation of the large volatility matrix. This is well documented in the statistics and econometrics literature even for the static large covariance matrix (Bickel and Levina 2008; Fan, Fan, and Lv 2008; Lam and Fan 2009; Rothman, Levina, and Zhu 2009). The additional challenge comes from the time-varying nature of a large volatility matrix. For a short and medium holding period (1 day or 1 week, say), the expected volatility matrix in the near future can be very different from the average of the expected volatility matrix over a long time horizon (the past 1 year, say). As a result, even if we know exactly the realized volatility matrix in the past, the bias can still be large. This calls for a stable and robust portfolio selection. The portfolio allocation under the gross-exposure constraint provides a needed solution. To reduce the bias of the forecasted expected volatility matrix, we need to shorten the learning period to better capture the dynamics of the time-varying volatility matrix, adapting better to the local volatility and correlation. But this is at the expense of a reduced sample size. The wide availability of high-frequency data provides sufficient amount of data for reliable estimation of the volatility matrix.

Recent years have seen dramatic developments in the study of high-frequency data in estimating integrated volatility. Statisticians and econometricians have been focusing on the interesting and challenging problem of volatility estimation in the presence of market microstructure noise and asynchronous tradings, which are the stylized features of high-frequency financial data. The progress is very impressive with a large literature. In particular, in the one-dimensional case when the focus is on estimation of integrated volatility, Aït-Sahalia, Mykland, and Zhang (2005) discussed a subsampling scheme; Zhang, Mykland, and Aït-Sahalia (2005) proposed a two-scale estimate which was extended and improved by Zhang (2006) to multiple scales; Fan and Wang (2007) separated jumps from diffusions in presence of market microstructural noise using a wavelet method; the robustness issues were addressed by Li and Mykland (2007); the realized kernel methods were proposed and thoroughly studied by Barndorff-Nielsen et al. (2009, 2011); Jacod et al. (2009) proposed a preaveraging approach to reduce the market microstructral noise; Xiu (2010) demonstrated that a simple quasi-likelihood method achieves the optimal rate of convergence for

estimating integrated volatility. For estimation of integrated co-variation, the nonsynchronized trading issue was first addressed by Hayashi and Yoshida (2005) in the absence of the microstructural noise; the kernel method with refresh time idea was first proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008); Zhang (2011) extended the two-scale method to study the integrated covariation using a previous tick method; Aït-Sahalia, Fan, and Xiu (2010) extended the quasi-maximum likelihood method; Kinnebrock, Podolskij, and Christensen (2010) extended the preaveraging technique. For the high-dimensional case, Wang and Zou (2010) estimated volatility matrix when sparsity condition is satisfied; Hautsch, Kyj, and Oomen (2009) studied a blocking and regularization approach; Tao et al. (2011) aggregated daily integrated volatility matrix via a factor model; Zheng and Li (2011) studied the empirical spectral distribution of the volatility matrix.

The aim of this article is to study the volatility matrix estimation using high-dimensional, high-frequency data from the perspective of financial engineering. Specifically, our main topics are how to extract the covariation information from high-frequency data for asset allocation and how effective they are. Two particular strategies are proposed for handling the nonsynchronized trading: "pairwise-refresh" and "all-refresh" schemes. The former retains many more data points and estimates covariance matrix element-by-element, which is often not positive semidefinite, whereas the latter retains far fewer data points and the resulting covariance matrix is more often positive semidefinite. As a result, the former has a better elementwise estimation error and is better in controlling risk approximation mentioned in the first paragraph of the introduction. However, the merits between the two methods are not that simple. In implementation, we need to project the estimate of covariance matrices onto the space of the semipositive definite matrices. The projections distort the accuracy of the elementwise estimation. As a result, the pairwise-refresh scheme does not have much more advantage than the all-refresh method, though the former is very easy to implement. However, both methods significantly outperform the methods based on low-frequency data, since they adapt better to the time-varying volatilities and correlations. The comparative advantage is more dramatic when there are rapid changes of the volatility matrix over time. This will be demonstrated in both simulation and empirical studies.

As mentioned before and demonstrated in Section 2, the accuracy of portfolio risk relative to the theoretically optimal portfolio is governed by the maximum elementwise estimation error. How does this error grow with the number of assets? Thanks to the concentration inequalities derived in this article, it grows only at the logarithmic order of the number of assets. This gives a theoretical endorsement why the portfolio selection problem is feasible for vast portfolios.

The article is organized as follows. Section 2 gives an overview of portfolio allocation using high-frequency data. Section 3 studies the volatility matrix estimation using high-frequency data from the perspective of asset allocation, where the analytical results are presented. How well our idea works in simulation and empirical studies can be found in Sections 4 and 5, respectively. Conclusions are given in Sec-

tion 6. Technical conditions and proofs are relegated to the Appendix.

2. CONSTRAINED PORTFOLIO OPTIMIZATION WITH HIGH-FREQUENCY DATA

2.1 Problem Setup

Consider a pool of p assets, with log-price processes $X_t^{(1)}$, $X_t^{(2)}, \ldots, X_t^{(p)}$. Denote by $\mathbf{X}_s = (X_s^{(1)}, \ldots, X_s^{(p)})^T$ the vector of the log-price processes at time s. Suppose they follow a diffusion process, namely,

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \mathbf{S}_t^{1/2} d\mathbf{W}_t \qquad (1)$$

where $\mathbf{W}_t$ is the vector of p-dimensional standard Brownian motions. The drift vector $\boldsymbol{\mu}_t$ and the instantaneous variance $\mathbf{S}_t$ can be stochastic processes and are assumed to be continuous.

A given portfolio with the allocation vector $\mathbf{w}$ at time t and a holding period τ has the log-return $\mathbf{w}^T \int_t^{t+\tau} dX_s$ with variance (risk)

$$R_{t,\tau}(\mathbf{w}) = \mathbf{w}^T \boldsymbol{\Sigma}_{t,\tau} \mathbf{w}, \qquad (2)$$

where $\mathbf{w}^T \mathbf{1} = 1$ and

$$\boldsymbol{\Sigma}_{t,\tau} = \int_t^{t+\tau} E_t \mathbf{S}_u du, \qquad (3)$$

with E_t denoting the conditional expectation given the history up to time t. Let w^+ be the proportion of long positions and w^- be the proportion of the short positions. Then, $\|\mathbf{w}\|_1 = w^+ + w^-$ is the gross exposure of the portfolio. To simplify the problem, following Jagannathan and Ma (2003) and Fan et al. (2011), we consider only the risk optimization problem. In practice, the expected return constraint can be replaced by the constraints of sectors or industries, to avoid unreliable estimates of the expected return vector. For a short-time horizon, the expected return is usually negligible. Following Fan et al. (2011), we consider the following risk optimization under gross-exposure constraints:

$$\min \mathbf{w}^T \boldsymbol{\Sigma}_{t,\tau} \mathbf{w}, \quad \text{s.t.} \|\mathbf{w}\|_1 \leq c \text{ and } \mathbf{w}^T \mathbf{1} = 1, \qquad (4)$$

where c is the total exposure allowed. Note that using $w^+ - w^- = 1$, problem (4) puts equivalently the constraint on the proportion of the short positions: $w^- \leq (c-1)/2$. As noted by Jagannathan and Ma (2003), the constrained optimization problem (4) is equivalent to unconstrained risk optimization with a regularized covariance matrix. Other methods of regularization are also possible to handle the noise-accumulation problem (e.g., the shrinkage method of Ledoit and Wolf (2004)).

The problem (4) involves the conditional expected volatility matrix (3) in the future. Unless we know exactly the dynamic of the volatility process, which is usually unknown, even if we observed the entire continuous paths up to the current time t. As a result, we rely on the approximation even with ideal data that we were able to observe the processes continuously without error. The typical approximation is

$$\tau^{-1} \boldsymbol{\Sigma}_{t,\tau} \approx h^{-1} \int_{t-h}^t \mathbf{S}_u du \qquad (5)$$

for an appropriate window width h and we estimate $\int_{t-h}^{t} \mathbf{S}_u du$ based on the historical data at the time interval $[t - h, t]$.

Approximation (5) holds reasonably well when τ and h are both small. This relies on the continuity assumptions: local time-varying volatility matrices are continuous in time. The approximation is also reasonable when both τ and h are large. This relies on the ergodicity assumption so that both quantities will be approximately $E\mathbf{S}_u$, when the stochastic volatility matrix $\mathbf{S}_u$ is ergodic. The approximation is not good when τ is small, whereas h is large as long as $\mathbf{S}_u$ is time varying, whether the stochastic volatility $\mathbf{S}_u$ is stationary or not. In other words, when the holding time horizon τ is short, as long as $\mathbf{S}_u$ is time varying, we can only use a short-time window $[t - h, t]$ to estimate $\mathbf{\Sigma}_{t,\tau}$. The recent arrivals of high-frequency data make this problem feasible.

The approximation error in (5) cannot usually be evaluated unless we have a specific parametric model on the stochastic volatility matrix $\mathbf{S}_u$. However, this is at the risk of model mis-specifications and nonparametric approach is usually preferred for high-frequency data. With p^2 elements approximated, which can be in the order of hundreds of thousands or millions, a natural question to ask is whether these errors accumulate and whether the result (risk) is stable. The gross-exposure constraint gives a stable solution to the problem as shown by Fan et al. (2011).

We would like to close this section by noting that formulation (4) is a one-period, not a multi-period, portfolio optimization problem.

2.2 Risk Approximations With Gross-Exposure Constraints

The utility of gross-exposure constraint can easily be seen through the following inequality. Let $\hat{\mathbf{\Sigma}}_{t,\tau}$ be an estimated covariance matrix and

$$\hat{R}_{t,\tau}(\mathbf{w}) = \mathbf{w}^T \hat{\mathbf{\Sigma}}_{t,\tau} \mathbf{w} \qquad (6)$$

be estimated risk of the portfolio. Then, for any portfolio with gross exposure $\|\mathbf{w}\|_1 \le c$, we have

$$\begin{aligned} |\hat{R}_{t,\tau}(\mathbf{w}) - R_{t,\tau}(\mathbf{w})| &\le \sum_{i=1}^{p}\sum_{j=1}^{p} |\hat{\sigma}_{i,j} - \sigma_{i,j}||w_i||w_j| \\ &\le |\mathbf{\Sigma}_{t,\tau} - \hat{\mathbf{\Sigma}}_{t,\tau}|_{\infty} \|\mathbf{w}\|_1^2 \\ &\le |\mathbf{\Sigma}_{t,\tau} - \hat{\mathbf{\Sigma}}_{t,\tau}|_{\infty} c^2, \end{aligned} \qquad (7)$$

where $\hat{\sigma}_{i,j}$ and $\sigma_{i,j}$ are respectively the (i, j) elements of $\hat{\mathbf{\Sigma}}_{t,\tau}$ and $\mathbf{\Sigma}_{t,\tau}$, and

$$|\mathbf{\Sigma}_{t,\tau} - \hat{\mathbf{\Sigma}}_{t,\tau}|_{\infty} = \max_{i,j} |\hat{\sigma}_{i,j} - \sigma_{i,j}|$$

is the maximum elementwise estimation error. Risk approximation (7) reveals that there is no large error accumulation effect when gross exposure c is moderate.

From now on, we drop the dependence of t and τ whenever there is no confusion. This facilitates the notation.

Fan et al. (2011) showed further that the risks of optimal portfolios are indeed close. Let

$$\begin{aligned} \mathbf{w}_{\text{opt}} &= \text{argmin}_{\mathbf{w}^T \mathbf{1}=1, \|\mathbf{w}\|_1 \le c} R(\mathbf{w}), \\ \hat{\mathbf{w}}_{\text{opt}} &= \text{argmin}_{\mathbf{w}^T \mathbf{1}=1, \|\mathbf{w}\|_1 \le c} \hat{R}(\mathbf{w}) \end{aligned} \qquad (8)$$

be, respectively, the theoretical (oracle) optimal allocation vector we want and the estimated optimal allocation vector we get. Then, $R(\mathbf{w}_{\text{opt}})$ is the theoretical minimum risk and $R(\hat{\mathbf{w}}_{\text{opt}})$ is the actual risk of our selected portfolio, whereas $\hat{R}(\hat{\mathbf{w}}_{\text{opt}})$ is our perceived risk, which is the quantity known to financial econometricians. They showed that

$$|R(\hat{\mathbf{w}}_{\text{opt}}) - R(\mathbf{w}_{\text{opt}})| \le 2a_p c^2, \qquad (9)$$
$$|R(\hat{\mathbf{w}}_{\text{opt}}) - \hat{R}(\hat{\mathbf{w}}_{\text{opt}})| \le a_p c^2, \qquad (10)$$
$$|R(\mathbf{w}_{\text{opt}}) - \hat{R}(\hat{\mathbf{w}}_{\text{opt}})| \le a_p c^2, \qquad (11)$$

with $a_p = |\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}|_{\infty}$, which usually grows slowly with the number of assets p. These reveal that the three relevant risks are in fact close as long as the gross-exposure parameter c is moderate and the maximum elementwise estimation error a_p is small.

The above risk approximations hold for any estimate of covariance matrix. It does not even require $\hat{\mathbf{\Sigma}}$ a semipositive definite matrix. This facilitates significantly the method of covariance matrix estimation. In particular, the elementwise estimation methods are allowed. In fact, since the approximation errors in (9), (10), and (11) are all controlled by the maximum elementwise estimation error, it can be advantageous to use elementwise estimation methods. This is particularly the case for the high-frequency data where trading is nonsynchronized. The synchronization can be done pairwisely or for all assets. The former retains much more data than the latter, as shown in the next section.

3. COVARIANCE MATRIX ESTIMATION USING HIGH-FREQUENCY DATA

3.1 All-Refresh Method and Pairwise-Refresh Method

Estimating high-dimensional volatility matrix using high-frequency data is a challenging task. One of the challenges is the nonsynchronicity of trading. Several synchronization schemes have been studied. The refresh time method was proposed by Barndorff-Nielsen et al. (2008) and the previous tick method was used by Zhang (2011). The former uses more efficiently the available data and will be used in this article.

The idea of refresh time is to wait until all assets are traded at least once at time v_1 (say) and then use the last price traded before or at v_1 of each asset as its price at time v_1. This yields one synchronized price vector at time v_1. The clock now starts again. Wait until all assets are traded at least once at time v_2 (say) and again use the previous tick price of each asset as its price at time v_2. This yields the second synchronized price vector at time v_2. Repeat the process until all available trading data are synchronized. Clearly, the process discards a large portion of the available trades. We will refer this synchorization scheme as the "all-refresh time." (The method is called all-refresh method for short.) Barndorff-Nielsen et al. (2008) advocated the kernel method to estimate integrated volatility matrix after synchronization; this can also be done by using other methods.

A more efficient method to use the available sample is the pairwise-refresh time scheme, which synchronizes the trading for each pair of assets separately. (The method is called pairwise-refresh method for short.) The pairwise-refresh scheme makes far more efficient use of the rich information in high-frequency

data, and enables us to estimate each element in the volatility matrix more precisely, which helps improve the efficiency of the selected portfolio. We will study the merits of these two methods. A third synchronization scheme is the blocking scheme of Hautsch et al. (2009) by grouping stocks with similar liquidities. The pairwise-refresh approach corresponds to the case with one-stock per group.

The pairwise estimation method allows us to use a wealth of univariate integrated volatility estimators such as those mentioned in the Introduction. For any given two assets with log-price processes $X_t^{(i)}$ and $X_t^{(j)}$, with pairwise-refresh times, the synchronized prices of $X_t^{(i)} + X_t^{(j)}$ and $X_t^{(i)} - X_t^{(j)}$ can be computed. With the univariate estimate of the integrated volatilities $< \widehat{X^{(i)} + X^{(j)}} >$ and $< \widehat{X^{(i)} - X^{(j)}} >$, the integrated covariation can be estimated as

$$\hat{\sigma}_{i,j} = \langle \widehat{X^{(i)}, X^{(j)}} \rangle = \left(\langle \widehat{X^{(i)} + X^{(j)}} \rangle - \langle \widehat{X^{(i)} - X^{(j)}} \rangle \right)/4. \quad (12)$$

In particular, the diagonal elements are estimated by the method itself. When the two-scale realized volatility (TSRV) (Zhang et al. 2005) is used, this results in the two-scale realized covariance (TSCV) estimate (Zhang 2011).

3.2 Pairwise-Refresh Method and TSCV

We now focus on the pairwise estimation method. To facilitate the notation, we reintroduce it.

We consider two log-price processes X and Y that satisfy

$$dX_t = \mu_t^{(X)}dt + \sigma_t^{(X)}dB_t^{(X)} \quad \text{and} \quad dY_t = \mu_t^{(Y)}dt + \sigma_t^{(Y)}dB_t^{(Y)}, \quad (13)$$

where $\text{cor}(B_t^{(X)}, B_t^{(Y)}) = \rho_t^{(X,Y)}$. For the two processes X and Y, consider the problem of estimating $\langle X, Y \rangle_T$ with $T = 1$. Denote by $\mathcal{T}_n$ the observation times of X and $\mathcal{S}_m$ the observation times of Y. Denote the elements in $\mathcal{T}_n$ and $\mathcal{S}_m$ by $\{\tau_{n,i}\}_{i=0}^n$ and $\{\theta_{m,i}\}_{i=0}^m$, respectively, in an ascending order ($\tau_{n,0}$ and $\theta_{m,0}$ are set to be 0). The actual log prices are not directly observable, but are observed with microstructure noises:

$$X_{\tau_{n,i}}^o = X_{\tau_{n,i}} + \epsilon_i^X, \quad \text{and} \quad Y_{\theta_{m,i}}^o = Y_{\theta_{m,i}} + \epsilon_i^Y, \quad (14)$$

where X^o and Y^o are the observed transaction prices in the logarithmic scale, and X and Y are the latent log prices govern by the stochastic dynamics (13). We assume that the microstructure noise ϵ_i^X and ϵ_i^Y processes are independent of the X and Y processes and that

$$\epsilon_i^X \overset{\text{i.i.d.}}{\sim} N(0, \eta_X^2) \quad \text{and} \quad \epsilon_i^Y \overset{\text{i.i.d.}}{\sim} N(0, \eta_Y^2). \quad (15)$$

Note that this assumption is mainly for the simplicity of presentation; as we can see from the proof, one can, for example, easily replace the identical Gaussian assumption with the not necessarily identically distributed (but are of the same variance) sub-Gaussian assumption without affecting our results.

The pairwise-refresh time $\mathcal{V} = \{v_0, v_1, \ldots, v_{\tilde{n}}\}$ can be obtained by setting $v_0 = 0$, and

$$v_i = \max \left\{ \min\{\tau \in \mathcal{T}_n : \tau > v_{i-1}\}, \ \min\{\theta \in \mathcal{S}_m : \theta > v_{i-1}\} \right\},$$

where $\tilde{n}$ is the total number of refresh times in the interval (0, 1]. The actual sample times for the two individual processes X and Y that correspond to the refresh times are

$$t_i = \max\{\tau \in \mathcal{T}_n : \tau \le v_i\} \quad \text{and} \quad s_i = \max\{\theta \in \mathcal{S}_m : \theta \le v_i\},$$

which are indeed the previous-tick measurement.

We study the property of the TSCV based on the *asynchronous* data:

$$\widehat{\langle X, Y \rangle}_1 = [X^o, Y^o]_1^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}[X^o, Y^o]_1^{(J)}, \quad (16)$$

where

$$[X^o, Y^o]_1^{(K)} = \frac{1}{K} \sum_{i=K}^{\tilde{n}} \left(X_{t_i}^o - X_{t_{i-K}}^o\right)\left(Y_{s_i}^o - Y_{s_{i-K}}^o\right)$$

and $\bar{n}_K = (\tilde{n} - K + 1)/K$. As discussed by Zhang (2011), the optimal choice of K has order $K = O(\tilde{n}^{2/3})$, and J can be taken to be a constant such as 1.

When either the microstructure error or the asynchronicity exists, the realized covariance is seriously biased. An asymptotic normality result in Zhang (2011) reveals that TSCV can simultaneously remove the bias due to the microstructure error and the bias due to the asynchronicity. However, this result is not adequate for our application to the vast volatility matrix estimation. To understand its impact on a_p, we need to establish the concentration inequality. In particular, for a sufficiently large $|x| = O((\log p)^\alpha)$, if

$$\max_{i,j} P\{\sqrt{n}|\sigma_{ij} - \hat{\sigma}_{ij}| > x\} < C_1 \exp\left(-C_2 x^{1/\alpha}\right), \quad (17)$$

for some positive constants C_1, C_2, and α, then

$$a_p = |\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}|_\infty = O_P\left(\frac{(\log p)^\alpha}{\sqrt{n}}\right). \quad (18)$$

We will show in the next section that the result indeed holds for some α which depends on the tail of the volatility process and n replaced by the minimum of the subsample size ($\bar{n}_K \sim (\tilde{n})^{\frac{1}{3}}$). Hence the impact of the number of assets is limited, only of the logarithmic order.

3.3 Concentration Inequalities

Inequality (17) requires the conditions on both diagonal elements and off-diagonal elements. Technically, they are treated differently. For the diagonal cases, the problem corresponds to the estimation of integrated volatility and there is no issue of asynchronicity. TSCV (16) reduces to TSRV (Zhang et al. 2005), which is explicitly given by

$$\widehat{\langle X, X \rangle}_1 = [X^o, X^o]_1^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}[X^o, X^o]_1^{(J)}, \quad (19)$$

where $[X^o, X^o]_1^{(K)} = \frac{1}{K} \sum_{i=K}^n (X_{t_i}^o - X_{t_{i-K}}^o)^2$ and $\bar{n}_K = (n - K + 1)/K$. As shown by Zhang et al. (2005), the optimal choice of K has order $K = O(n^{2/3})$ and J can be taken to be a constant such as 1.

To facilitate the reading, we relegate the technical conditions and proofs to the Appendix. The following two results establish the concentration inequalities for the integrated volatility and integrated covariation.

Theorem 1. Let X process be as in (13), and n be the total number of observations for the X process during the time interval (0,1]. Under Conditions 1–4 in the Appendix

(A) if $\sigma_t^{(X)} \le C_\sigma < \infty$ for all $t \in [0, 1]$, then for all large n, for $x \in [0, cn^{1/6}]$,

$$P\left\{n^{1/6}|\widehat{\langle X, X \rangle}_1 - \int_0^1 \sigma_t^{(X)2}dt| > x\right\} \le 4\exp\{-Cx^2\}$$

for positive constants c and C. A set of candidate values for c and C are given in (A.25).

(B) if the tail behavior of $\sigma_t^{(X)}$ can be described as

$$P\left\{\sup_{0 \leq t \leq 1} \sigma_t^{(X)} \geq C_\sigma\right\} \leq k_\sigma \exp\left\{-aC_\sigma^b\right\}, \text{ for any } C_\sigma > 0$$

with positive constants k_σ, a and b, then for all large n, for $x \in [0, cn^{\frac{4+b}{6b}}]$,

$$P\left\{n^{1/6}|\widehat{\langle X, X\rangle}_1 - \int_0^1 \sigma_t^{(X)2} dt| > x\right\}$$
$$\leq (4 + k_\sigma) \exp\{-C \cdot x^{\frac{2b}{4+b}}\}.$$

A set of candidate values for c and C are given in (A.26).

Theorem 2. Let X and Y be as in (13), and $\tilde{n}$ be the total number of refresh times for the processes X and Y during the time interval $(0,1]$. Under Conditions 1–5 in the Appendix,

(A) if $\sigma_t^{(i)} \leq C_\sigma < \infty$ for all $t \in [0, 1]$ and $i = X$ and Y, then for all large $\tilde{n}$, for $x \in [0, c\tilde{n}^{1/6}]$,

$$P\left\{\tilde{n}^{1/6}|\widehat{\langle X, Y\rangle}_1 - \int_0^1 \sigma_t^{(X)}\sigma_t^{(Y)}\rho_t^{(X,Y)} dt| > x\right\}$$
$$\leq 8 \exp\{-Cx^2\}$$

for positive constants c and C. A set of candidate values for c and C are given in (A.29).

(B) if the tail behavior of $\sigma_t^{(i)}$ for $i = X$ or Y satisfy

$$P\left\{\sup_{0 \leq t \leq 1} \sigma_t^{(i)} \geq C_\sigma\right\} \leq k_\sigma \exp\left\{-aC_\sigma^b\right\}, \text{ for any } C_\sigma > 0$$

with positive constants k_σ, a, and b, then for all large $\tilde{n}$, for $x \in [0, c\tilde{n}^{\frac{4+b}{6b}}]$,

$$P\left\{\tilde{n}^{1/6}|\widehat{\langle X, Y\rangle}_1 - \int_0^1 \sigma_t^{(X)}\sigma_t^{(Y)}\rho_t^{(X,Y)} dt| > x\right\}$$
$$\leq (8 + 2k_\sigma) \exp\left\{-Cx^{\frac{2b}{4+b}}\right\}.$$

A set of candidate values for c and C are given in (A.30).

3.4 Error Rates on Risk Approximations

Having had the above concentration inequalities, we can now readily give an upper bound of the risk approximations. Consider the p log-price processes as in Section 2.1. Suppose the processes are observed with the market microstructure noises. Let $\tilde{n}^{(i,j)}$ be the observation frequency obtained by the pairwise-refresh method for two processes $X^{(i)}$ and $X^{(j)}$ and $\tilde{n}_*$ be the observation frequency obtained by the all-refresh method. Clearly, $\tilde{n}^{(i,j)}$ is typically much larger than $\tilde{n}_*$. Hence, most elements are estimated more accurately using the pairwise-refresh method than using the all-refresh method. On the other hand, for less liquidly traded pairs, its observation frequency of pairwise-refresh time cannot be an order of magnitude larger than $\tilde{n}_*$.

Using (18), an application to Theorems 1 and 2 to each element in the estimated integrated covariance matrix yields

$$a_p^{\text{pairwise-refresh}} = |\hat{\Sigma}^{\text{pairwise}} - \Sigma|_\infty = O_P\left(\frac{(\log p)^\alpha}{\tilde{n}_{\min}^{1/6}}\right), \quad (20)$$

where α is $\frac{1}{2}$ when the volatility processes are bounded and is a constant depending on the tail behavior of the volatility

processes when they are unbounded, and $\tilde{n}_{\min} = \min_{i,j} \tilde{n}^{(i,j)}$ is the minimum number of observations of the pairwise-refresh time.

Note that based on our proofs which do not rely on any particular properties of pairwise-refresh times, our results of Theorems 1 and 2 are applicable to all-refresh method as well, with the observation frequency of the pairwise-refresh times replaced by that of the all-refresh times. Hence, using the all-refresh time scheme, we have

$$a_p^{\text{all-refresh}} = |\hat{\Sigma}^{\text{all-refresh}} - \Sigma|_\infty = O_P\left(\frac{(\log p)^\alpha}{\tilde{n}_*^{1/6}}\right), \quad (21)$$

with the same α as above. Clearly, $\tilde{n}_{\min}$ is larger than $\tilde{n}_*$. Hence, the pairwise-refresh method gives a somewhat more accurate estimate in terms of the maximum elementwise estimation error.

3.5 Projections of Estimated Volatility Matrices

The risk approximations (9)–(11) hold for any solutions to (8) whether the matrix $\hat{\Sigma}$ is positive semidefinite or not. However, convex optimization algorithms typically require the positive semidefiniteness of the matrix $\hat{\Sigma}$. Yet, the estimates based on the elementwise estimation sometimes cannot all satisfy this and even the ones from all-refresh method can have the same problem when TSCV is applied. This leads to the issue of how to project a symmetric matrix onto the space of positive semidefinite matrices.

There are two intuitive methods for projecting a $p \times p$ symmetric matrix $\mathbf{A}$ onto the space of positive semidefinite matrices. Consider the singular value decomposition: $\mathbf{A} = \mathbf{\Gamma}^T \text{diag}(\lambda_1, \ldots, \lambda_p)\mathbf{\Gamma}$, where $\mathbf{\Gamma}$ is an orthogonal matrix, consisting of p eigenvectors. The two intuitive appealing projection methods are

$$\mathbf{A}_1^+ = \mathbf{\Gamma}^T \text{diag}(\lambda_1^+, \ldots, \lambda_n^+)\mathbf{\Gamma}, \quad (22)$$

where λ_j^+ is the positive part of λ_j, and

$$\mathbf{A}_2^+ = (\mathbf{A} + \lambda_{\min}^- I_p)/(1 + \lambda_{\min}^-), \quad (23)$$

where $\lambda_{\min}^-$ is the negative part of the minimum eigenvalue. For both projection methods, the eigenvectors remain the same as those of $\mathbf{A}$. When $\mathbf{A}$ is positive semidefinite matrix, we have obviously that $\mathbf{A}_1 = \mathbf{A}_2 = \mathbf{A}$.

In applications, we apply the above transformations to the estimated correlation matrix $\mathbf{A}$ rather than directly to the volatility matrix estimate $\hat{\Sigma}$. The correlation matrix $\mathbf{A}$ has diagonal elements of 1. The resulting matrix under projection method (23) apparently still satisfies this property, whereas the one under projection method (22) does not. As a result, projection method (23) keeps the integrated volatility of each asset intact.

In our initial simulation and empirical studies, we applied both projections. It turns out that there is no significant difference between the two projection methods in terms of results. We decided to apply only projection (23) in all numerical studies.

3.6 Comparisons Between Pairwise- and All-Refresh Methods

The pairwise-refresh method keeps far richer information in the high-frequency data than the all-refresh method. Thus, it is

expected to estimate each element more precisely. Yet, the estimated correlation matrix is typically not positive semidefinite. As a result, projection (23) can distort the accuracy of element-wise estimation. On the other hand, the all-refresh method more often gives positive semidefinite estimates. Therefore, projection (23) has less impact on the all-refresh method than on the pairwise-refresh method.

Risk approximations (9)–(11) are only the upper bounds. The upper bounds are controlled by a_p, which has rates of convergence govern by (20) and (21). While the average number of observations of pairwise-refresh time is far larger than the number of observations $\tilde{n}_*$ of the all-refresh time, the minimum number of observations of the pairwise-refresh time $\tilde{n}_{\min}$ is not much larger than $\tilde{n}_*$. Therefore, the upper bounds (20) and (21) are approximately of the same level. This together with the distortion due to projection do not leave much advantage for the pairwise-refresh method.

4. SIMULATION STUDIES

In this section, we simulate the market trading data using a reasonable stochastic model. As the latent prices and dynamics of simulations are known, our study on the risk profile is facilitated. In particular, we would like to verify our theoretical results and to quantify the finite sample behaviors.

In this section and the next, the risk refers to the standard deviation of portfolio's returns. To avoid ambiguity, we call $\sqrt{R(\mathbf{w}_{\text{opt}})}$ the theoretical optimal risk or oracle risk, $\sqrt{\hat{R}(\hat{\mathbf{w}}_{\text{opt}})}$ the perceived optimal risk, and $\sqrt{R(\hat{\mathbf{w}}_{\text{opt}})}$ the actual risk of the perceived optimal allocation.

4.1 Design of Simulations

A slightly modified version of the simulation model in Barndorff-Nielsen et al. (2008) is used to generate the latent price processes of p traded assets. It is a multivariate factor model with stochastic volatilities. Specifically, the latent log prices $X_t^{(i)}$ follow

$$dX_t^{(i)} = \mu^{(i)}dt + \rho^{(i)}\sigma_t^{(i)}dB_t^{(i)} + \sqrt{1-(\rho^{(i)})^2}\sigma_t^{(i)}dW_t + \nu^{(i)}dZ_t,$$
$$i = 1, \ldots, p, \tag{24}$$

where the elements of B, W, and Z are independent standard Brownian motions. The spot volatility obeys the independent Ornstein–Uhlenbeck processes:

$$d\varrho_t^{(i)} = \alpha^{(i)}(\beta_0^{(i)} - \varrho_t^{(i)})dt + \beta_1^{(i)}dU_t^{(i)}, \tag{25}$$

where $\varrho_t^{(i)} = \log \sigma_t^{(i)}$ and $U_t^{(i)}$ is an independent Brownian motion.

The number of assets p is taken to be 50. Slightly modified from Barndorff-Nielsen et al. (2008), the parameters are set to be $(\mu^{(i)}, \beta_0^{(i)}, \beta_1^{(i)}, \alpha^{(i)}, \rho^{(i)}) = (0.03x_1^{(i)}, -x_2^{(i)}, 0.75x_3^{(i)}, -1/40x_4^{(i)}, -0.7)$ and $\nu^{(i)} = \exp(\beta_0^{(i)})$, where $x_j^{(i)}$ is an independent realization from the uniform distribution on $[0.7, 1.3]$. The parameters are kept fixed in the simulations.

Model (24) is used to generate the latent log-price values with initial values $X_0^{(i)} = 1$ (log price) and $\varrho_0^{(i)}$ from its stationary distribution. The Euler scheme is used to generate latent price at the frequency of once per second. To account for the market

microstructure noise, the Gaussian noises $\varepsilon_t^{(i)} \overset{\text{i.i.d.}}{\sim} N(0, \omega^2)$ with $\omega = 0.0005$ are added. Therefore, like (14), the observed log prices are $X_t^{o(i)} = X_t^{(i)} + \varepsilon_t^{(i)}$.

To model the nonsynchronicity, p independent Poisson processes with intensity parameters $\lambda_1, \lambda_2, \ldots, \lambda_p$ are used to simulate the trading times of the assets. Motivated by the U.S. equity trading dataset (the total number of seconds in a common trading day of the U.S. equity is 23, 400), we set the trading intensity parameters λ_i's to be $0.02i \times 23400$ for $i = 1, 2, \ldots, 50$, meaning that the average numbers of trading times for each asset are spread out in the arithmetic sequence of the interval [468, 23,400].

4.2 An Oracle Investment Strategy and Risk Assessment

An oracle investment strategy is usually a decent benchmark for other portfolio strategies to compare with. There are several oracle strategies. The one we choose is to make portfolio allocation based on the covariance matrix estimated using latent prices at the finest grid (one per second). Latent prices are the noise-free prices of each asset at every time points (one per second), which are unobservable in practice and are available to us only in the simulation. Therefore, for each asset, there are 23, 400 latent prices in a normal trading day. We will refer to the investment strategy based on the latent prices as the oracle or latent strategy. This strategy is not available for the empirical studies.

The assessment of risk is based on the high-frequency data. For a given portfolio strategy, its risk is computed based on the latent prices at every 15 minutes for the simulation studies, whereas for the empirical studies, the observed prices at every 15 minutes are used to assess its risk. This mitigates the influence of the microstructure noises. For the empirical study, we do not hold positions overnight therefore are immune to the overnight price jumps (we will discuss the details in Section 5).

4.3 Out-of-Sample Optimal Allocation

One of the main purposes of this article is to investigate the comparative advantage of the high-frequency-based methods against the low-frequency-based methods, especially in the context of portfolio investment. Hence, it is essential for us to run the following out-of-sample investment strategy test which includes both the high-frequency- and low-frequency-based approaches. Moreover, since in the empirical studies, we do not know the latent asset prices, the out-of-sample test should be designed so that it can also be conducted in the empirical studies.

We simulate the prices of 50 traded assets as described in Section 4.1 for the duration of 200 trading days (numbered as day 1, day 2, . . . , day 200) and record all the tick-by-tick trading times and trading prices of the assets.

We start investing 1 unit of capital into the pool of assets with low-frequency- and high-frequency-based strategies from day 101 (the portfolios are bought at the opening of day 101). For the low-frequency strategy, we use the previous 100 trading days' daily closing prices to compute the sample covariance matrix and to make the portfolio allocation accordingly with the gross-exposure constraints. For the high-frequency strategies, we use the previous $h = 10$ trading days' tick-by-tick trading data. For

the all-refresh strategy, we use all-refresh time to synchronize the trades of the assets before applying TSCV to estimate the integrated volatility matrix and make the portfolio allocation, while for the pairwise-refresh high-frequency strategy, we use pairwise-refresh times to synchronize each pair of assets and apply TSCV to estimate the integrated covariance for the corresponding pair. With projection technique (23), the resulting TSCV integrated volatility matrix can always be transformed to a positive semidefinite matrix which facilitates the optimization.

We run two investment plans. In the first plan, the portfolio is held for $\tau = 1$ trading day before we reestimate the covariation structure and adjust the portfolio weights accordingly. The second plan is the same as the first one except for the fact that the portfolio is held for $\tau = 5$ trading days before rebalancing.

In the investment horizon (which is from day 101 to day 200 in this case), we record the 15-minute portfolio returns based on the latent prices of the assets, the variation of the portfolio weights across 50 assets, and other relevant characteristics. While it appears that 100 trading days is short, calculating 15-minute returns increases the size of the relevant data for computing the risk by a factor of 26.

We study those portfolio features for a whole range of gross-exposure constraint c from $c = 1$, which stands for the no-short-sale portfolio strategy, to $c = 3$. This is usually the relevant range of gross exposure for investment purpose.

The standard deviations and other characteristics of the strategy for $\tau = 1$ are presented in Table 1 (the case $\tau = 5$ gives similar comparisons and hence is omitted). The standard deviations, which are calculated based on the 15-minute returns as mentioned before, represent the actual risks of the strategy. As we only optimize the risk profile, we should not look significantly on the returns of the optimal portfolios. They cannot even be estimated with good accuracy over such a short investment horizon. Figures 1 and 2 provide graphical details to these characteristics for both $\tau = 1$ and $\tau = 5$.

Table 1. The out-of-sample performance of daily-rebalanced optimal portfolios with gross-exposure constraint

Methods	Std Dev %	Max Weight	Min Weight	No. of Long	No. of Short
Low-frequency sample covariance matrix estimator					
$c = 1$ (No short)	16.69	0.19	−0.00	13	0
$c = 2$	16.44	0.14	−0.05	28.5	20
$c = 3$	16.45	0.14	−0.05	28.5	20
High-frequency all-refresh TSCV covariance matrix estimator					
$c = 1$ (No short)	16.08	0.20	−0.00	15	0
$c = 2$	14.44	0.14	−0.05	30	19
$c = 3$	14.44	0.14	−0.05	30	19
High-frequency pairwise-refresh TSCV covariance matrix estimator					
$c = 1$ (No short)	15.34	0.18	−0.00	15	0
$c = 2$	12.72	0.13	−0.03	31	18
$c = 3$	12.72	0.13	−0.03	31	18

NOTE: We simulate one trial of intra-day trading data for 50 assets, make portfolio allocations for 100 trading days, and rebalance daily. The standard deviations and other characteristics of these portfolios are recorded. All the characteristics are annualized (Max Weight, median of maximum weights; Min Weight, median of minimum weights; No. of Long, median of numbers of long positions whose weights exceed 0.001; No. of Short, median of numbers of short positions whose absolute weights exceed 0.001).

From Table 1 and Figures 1 and 2, we see that for both holding lengths $\tau = 1$ and $\tau = 5$, the all-refresh TSCV and pairwise-refresh TSCV approaches outperform significantly the low-frequency one in terms of risk profile for the whole range of the gross-exposure constraints. This supports our theoretical results and intuitions. First, the shorter estimation window allows these two high-frequency approaches to deliver consistently better results than the low-frequency one. Second, the pairwise-refresh method outperforms the all-refresh method, as expected. Finally, both the low-frequency strategy and the high-frequency strategies outperform significantly the equal-weight portfolio (see Figures 1 and 2).

All the risk curves attain their minimum around $c = 1.2$ (see Figures 1 and 2), which meets our expectation again, since that must be the point where the marginal increase in estimation error outpaces the marginal decrease in specification error. This, coupled with the result we get in Empirical Studies section, will give us some guidelines about what gross-exposure constraint to use in investment practice.

In terms of portfolio weights, neither the low-frequency nor the high-frequency optimal no-short-sale portfolios are well diversified with all approaches assigning a concentrated weight of around 20% to one individual asset. Their portfolio risks can be improved by relaxing the gross-exposure constraint (see Figures 1 and 2).

5. EMPIRICAL STUDIES

Risk minimization problem (6) has important applications in asset allocation. We demonstrate its application in the stock portfolio investment in the 30 Dow Jones Industrial Average (30 DJIA) constituent stocks.

To make asset allocation, we use the high-frequency data of the 30 DJIA stocks from January 1, 2008, to September 30, 2008. These stocks are highly liquid. The period covers the birth of financial crisis in 2008.

At the end of each holding period of $\tau = 1$ or $\tau = 5$ trading days in the investment period (from May 27, 2008, to September 30, 2008), the covariance of the 30 stocks is estimated according to the different estimators. They are the sample covariance of the previous 100 trading days' daily return data (low-frequency), the all-refresh TSCV estimator of the previous 10 trading days, and the pairwise-refresh TSCV estimator of the previous 10 trading days. These estimated covariance matrices are used to construct optimal portfolios with a range of exposure constraints. For $\tau = 5$, we do not count the overnight risks of the portfolio. The reason is that the overnight price jumps are often due to the arrival of news and are irrelevant of the topics of our study. The standard deviations and other characteristics of these portfolio returns for $\tau = 1$ are presented in Table 2 together with the standard deviation of an equally weighted portfolio of the 30 DJIA stocks rebalanced daily. The standard deviations are for the 15 minutes returns, which represent the actual risks. Figures 3 and 4 provide the graphical details to these characteristics for both $\tau = 1$ and $\tau = 5$.

Table 2 and Figures 3 and 4 reveal that in terms of the portfolio's actual risk, the all-refresh TSCV and pairwise-refresh TSCV strategies perform at least as well as the low-frequency-based strategy when the gross exposure is small and outperform the latter significantly when the gross exposure
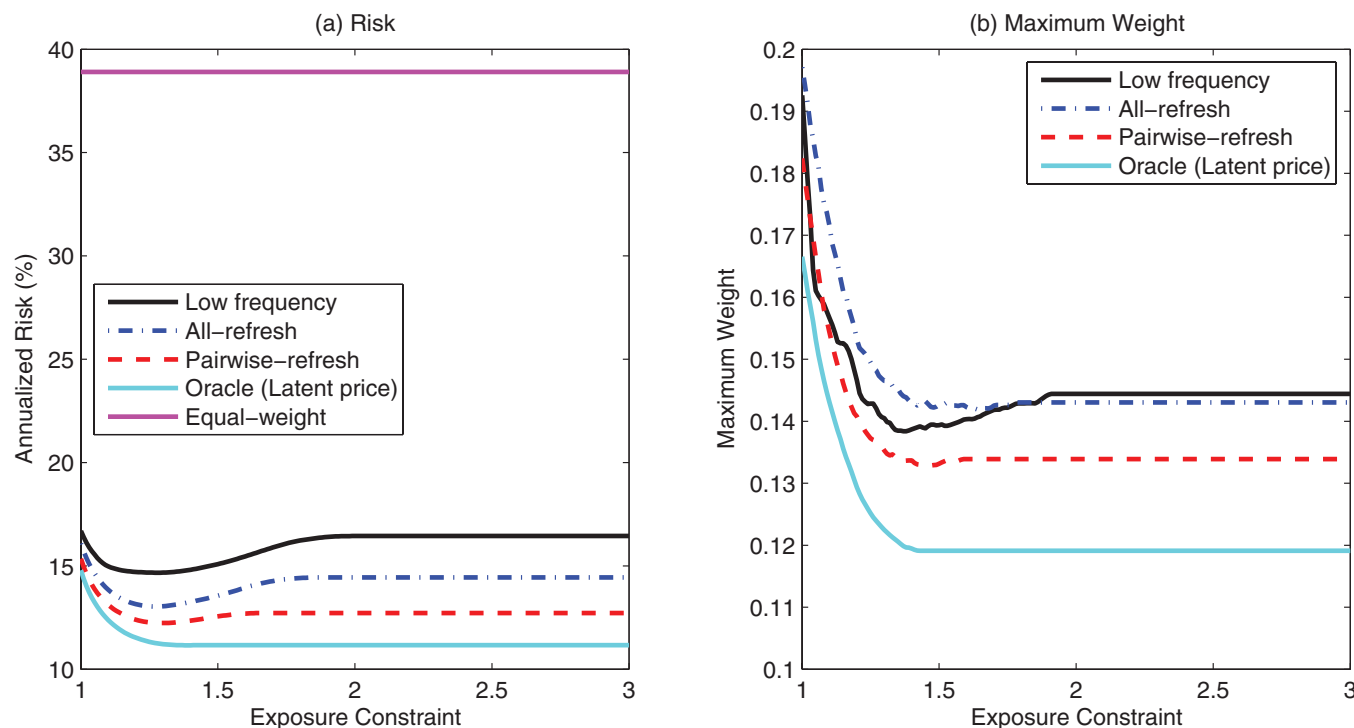
Figure 1. Out-of-sample performance of daily-rebalanced optimal portfolios based on high-frequency and low-frequency estimation of the integrated covariance matrix. (a) Annualized risk of portfolios. (b) Maximum weight of allocations. (The online version of this figure is in color.)

is large. Both facts support our theoretical results and intuitions. Given 10 times the length of covariance estimation window, the low-frequency approach still cannot perform better than the high-frequency TSCV approaches, which affirms our belief that the high-frequency TSCV approaches can significantly shorten

the necessary covariance estimation window and capture better the short-term time-varying covariation structure (or the "local" covariance). These results, together with the ones presented in the Simulation Studies section, lend strong support to the above statement.
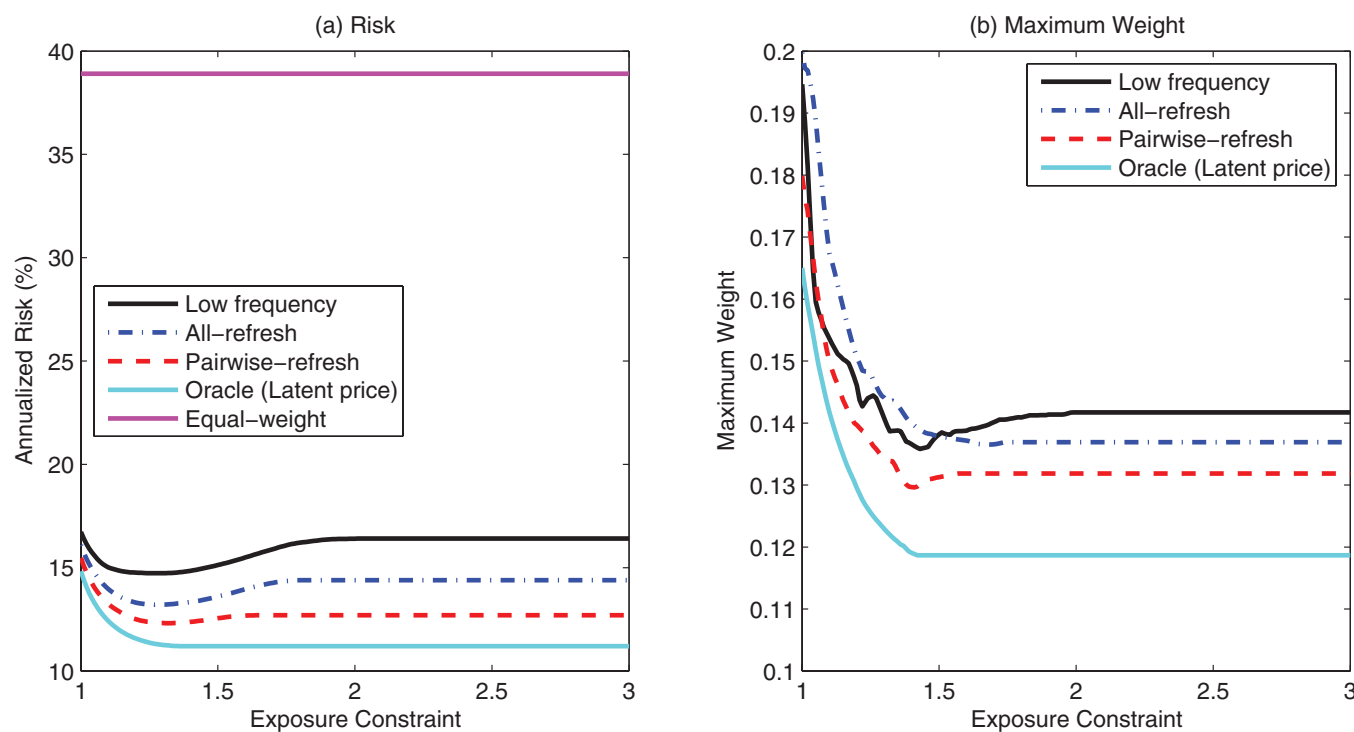

Figure 2. Out-of-sample performance of optimal portfolios based on high-frequency and low-frequency estimation of the integrated covariance matrix with holding period $\tau = 5$. (a) Annualized risk of portfolios, (b) maximum weight of allocations. (The online version of this figure is in color.)

Table 2. The out-of-sample performance of daily-rebalanced optimal portfolios of the 30 DJIA stocks

Methods	Std Dev %	Max Weight	Min Weight	No. of Long	No. of Short
Low-frequency sample covariance matrix estimator					
$c = 1$ (No short)	12.73	0.50	−0.00	8	0
$c = 2$	14.27	0.44	−0.12	16	10
$c = 3$	15.12	0.45	−0.18	18	12
High-frequency all-refresh TSCV covariance matrix estimator					
$c = 1$ (No short)	12.55	0.40	−0.00	8	0
$c = 2$	12.36	0.36	−0.10	17	12
$c = 3$	12.50	0.36	−0.10	17	12
High-frequency pairwise-refresh TSCV covariance matrix estimator					
$c = 1$ (No short)	12.54	0.39	−0.00	9	0
$c = 2$	12.23	0.35	−0.08	17	12
$c = 3$	12.34	0.35	−0.08	17	12
Unmanaged index					
30 DJIA equally weighted	22.12				

As the gross-exposure constraint increases, the portfolio risk of the low-frequency approach increases drastically relative to the ones of the high-frequency TSCV approaches. The reason could be a combination of the fact that the low-frequency approach does not produce a well-conditioned estimated covariance due to the lack of data and the fact that the low-frequency approach can only attain the long-run covariation but cannot capture well the "local" covariance dynamics. The portfolio risk of the high-frequency TSCV approaches increased only moderately as the gross-exposure constraint increases. From financial practitioner's standpoint, that is also one of the comparative advantages of high-frequency TSCV approaches, which means that investors do not need to be much concerned about the choice of the gross-exposure constraint while using the high-frequency TSCV approaches.

It can be seen that both the low-frequency and high-frequency optimal no-short-sale portfolios are not diversified enough. Their risk profiles can be improved by relaxing the gross-exposure constraint to around $c = 1.2$, that is, 10% short positions and 110% long positions are allowed. The no-short-sale portfolios under all approaches have the maximum portfolio weight of 22%–50%. As the gross-exposure constraint relaxes, the pairwise-refresh TSCV approach has its maximum weight reaching the smallest value around 30%–34%, while the low-frequency approach goes down to only around 40%. This is another comparative advantage of the high-frequency approach in practice, as a portfolio with less weight concentration is typically considered more preferable.

Another interesting fact is that the equally weighted daily-rebalanced portfolio of the 30 DJIA stocks carries an annualized return of only −10% while DJIA went down 13.5% during the same period (May 27, 2008, to Sep 30, 2008), giving an annualized return of −38.3%. The cause of the difference is that we intentionally avoided holding portfolios overnight, hence the portfolios are not affected by the overnight price jumps. In the turbulent financial market of May to September 2008, our portfolio strategies are not affected by the numerous sizeable downward jumps. Those jumps are mainly caused by the news of distressed economy and corporations. The moves could deviate far from what the previously held covariation structure dictates.

Figure 3. Out-of-sample performance of daily-rebalanced optimal portfolios for 30 DJIA constituent stocks with investment period from May 27, 2008, to September 30, 2008 (89 trading days). (a) Annualized risk of portfolios. (b) Maximum weight of allocations. (The online version of this figure is in color.)
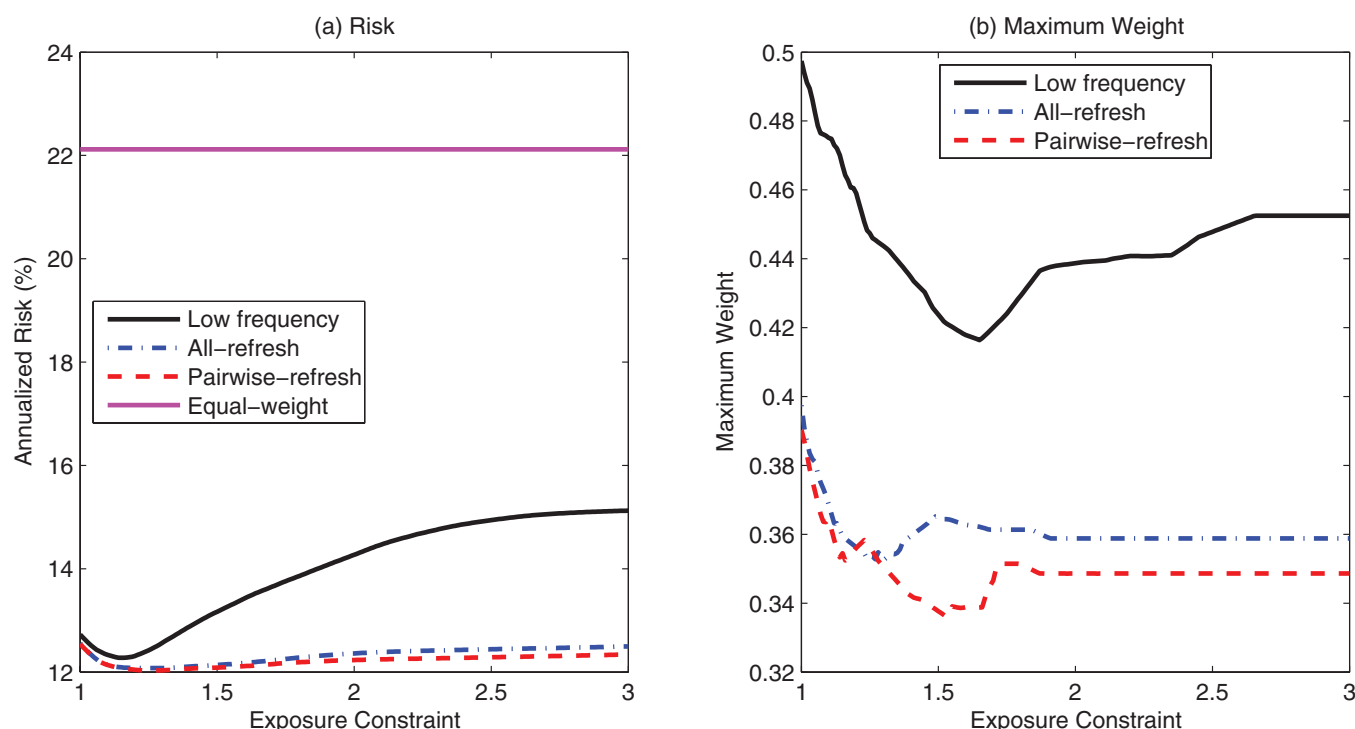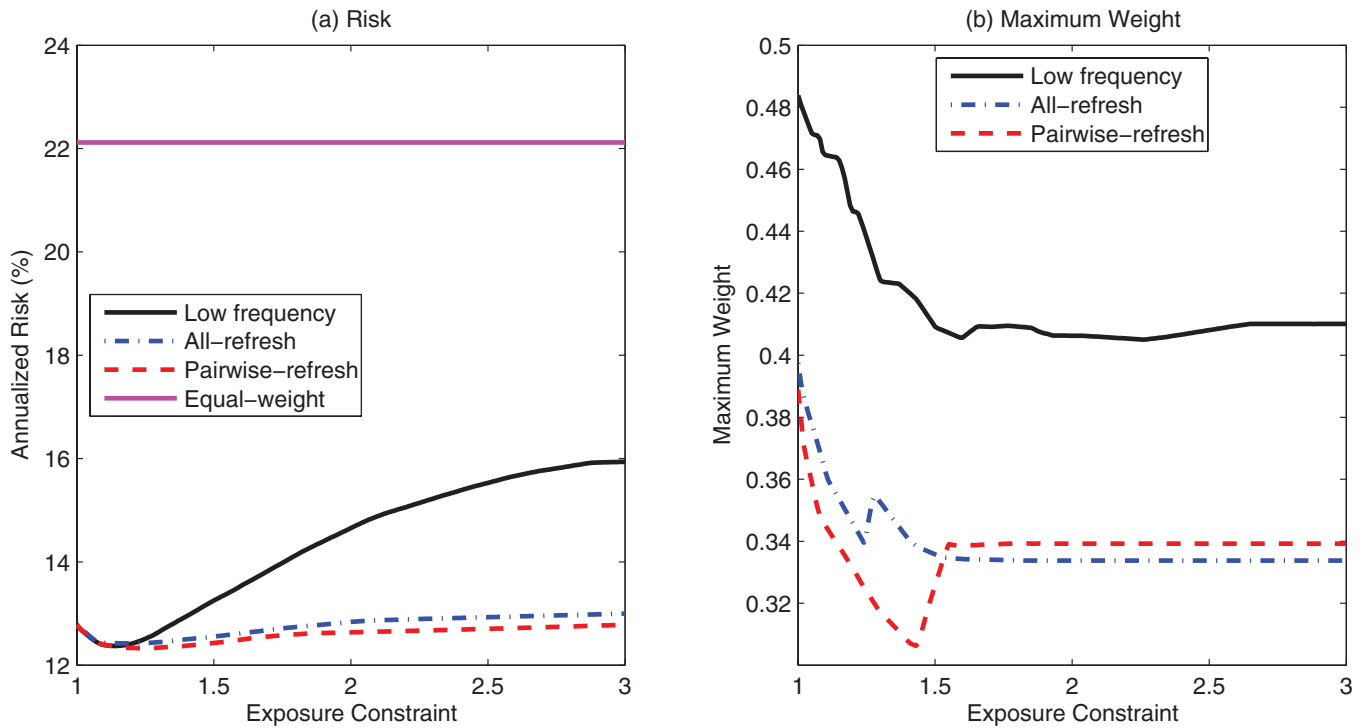
Figure 4. Out-of-sample performance of 5-day-rebalanced optimal portfolios for 30 DJIA constituent stocks with investment period from May 27, 2008, to September 30, 2008 (89 trading days). (a) Annualized risk of portfolios. (b) Maximum weight of allocations. (The online version of this figure is in color.)

6. CONCLUSION

We advocate the portfolio selection with gross-exposure constraint (Fan et al. 2011). It is less sensitive to the error of co-variance estimation and mitigates the noise accumulation. The out-of-sample portfolio performance depends on the expected volatility in the holding period. It is at best approximated and the gross-exposure constraints help reducing the error accumulation in the approximations.

Two approaches are proposed for the use of high-frequency data to estimate the integrated covariance: "all-refresh" and "pairwise-refresh" methods. The latter retains far more data on average and hence estimates more precisely element-by-element. Yet, the pairwise-refresh estimates are often not positive semidefinite and projections are needed for the convex optimization algorithms. The projection distorts somewhat the performance of the pairwise-refresh strategies. New optimization algorithms need to be developed in order to take full advantage of pairwise-refresh. Further investigations on the relative merits of "pairwise-refresh," "blocking approach," and "all-refresh" are needed.

The use of high-frequency financial data increases significantly the available sample size for volatility estimation, and hence shortens the time window for estimation, adapts better to local covariations. Our theoretical observations are supported by the empirical studies and simulations, in which we demonstrate convincingly that the high-frequency-based strategies outperform the low-frequency-based one in general.

With the gross-exposure constraint, the impact of the size of the candidate pool for portfolio allocation is limited. We derive the concentration inequalities to demonstrate this theoretically. Simulation and empirical studies also lend further support to it.

APPENDIX: CONDITIONS AND PROOFS

Conditions

We derive our theoretical results under the following conditions. For simplicity, we state the conditions for integrated covariation (Theorem 2). The conditions for integrated volatility (Theorem 1) are simply the ones with $Y = X$.

Condition 1. The drift processes are such that $\mu_t^{(X)} = \mu_t^{(Y)} = 0$ for all $t \in [0, 1]$.

Condition 2. $\sigma_t^{(i)}$, $i = X, Y$ are continuous stochastic processes which are either bounded by $0 < C_\sigma < \infty$, or such that the tail behavior can be described by

$$P\left\{\sup_{0 \le t \le 1} \sigma_t^{(i)} \ge C_\sigma\right\} \le k_\sigma \exp\left\{-aC_\sigma^b\right\}, \quad \text{for any } C_\sigma > 0,$$

with positive constants k_σ, a, and b.

Condition 3. The observation times are independent with the X and Y processes. The synchronized observation times for the X and Y processes satisfy $\sup_{1 \le j \le \tilde{n}} \tilde{n} \cdot (v_j - v_{j-1}) \le C_\Delta \le \infty$, where C_Δ is a nonrandom constant, $\tilde{n}$ is the observation frequency, and $\mathcal{V} = \{v_0, v_1, \ldots, v_{\tilde{n}}\}$ is the set of refresh times of the processes X and Y.

Condition 4. For the TSCV parameters, we consider the case when $J = 1$ ($\bar{n}_J = \tilde{n}$) and $\bar{n}_K = O(\tilde{n}^{1/3})$ such that $\frac{1}{2} \cdot \tilde{n}^{1/3} \le \bar{n}_K \le 2 \cdot \tilde{n}^{1/3}$.

Condition 5. The processes ϵ^X and ϵ^Y are independent.

Remark 1. Conditions 1 and 4 are imposed for simplicity. They can be removed at the expenses of lengthier proofs. For a short horizon and high-frequency, whether Condition 1 holds or not has little impact on the investment. For estimating integrated volatility, the synchronized time becomes observation time $\{\tau_{n,j}\}$ and Conditions 3 and 4

become

$$\sup_{1 \le j \le n} n \cdot (\tau_{n,j} - \tau_{n,j-1}) \le C_\Delta < \infty \qquad (A.1)$$

and $\frac{1}{2} \cdot n^{1/3} \le \bar{n}_K \le 2 \cdot n^{1/3}$.

Remark 2. Note that indeed, in the derivations of the Theorems 1 and 2, $\sigma_t^{(X)}$ and $\sigma_t^{(Y)}$ are only required to be càdlàg. In other words, the continuity assumption is not needed for the concentration inequalities to hold. The continuity of the volatility processes is only needed in approximation (5) in our study, it can be removed or relaxed for other applications of the concentration inequalities.

Lemmas

We need the following three lemmas for the proofs of Theorems 1 and 2. In particular, Lemma 2 is exponential type of inequality for any dependent random variables that have a finite moment generating function. It is useful for many statistical learning problems. Lemma 3 is a concentration inequality for the realized volatility based on discretely observed latent process.

Lemma 1. When $Z \sim N(0,1)$, for any $|\theta| \le \frac{1}{4}$, $E \exp\{\theta(Z^2 - 1)\} \le \exp(2\theta^2)$.

Proof. Using the moment generating function of $Z^2 \sim \chi_1^2$, we have

$$E \exp\{\theta(Z^2 - 1)\} = \exp\left\{ -\frac{1}{2} \log(1 - 2\theta) - \theta \right\}.$$

Let $g(x) = \log(1-x) + x + x^2$ with $|x| \le 1/2$. Then, $g'(x) = x(1-2x)/(1-x)$ is nonnegative when $x \in [0, 1/2]$ and negative when $x \in [-1/2, 0)$. In other words, $g(x)$ has a minimum at point 0, namely $g(x) \ge 0$ for $|x| \le 1/2$. Consequently, for $|\theta| \le 1/4$, $\log(1 - 2\theta) \ge -2\theta - (2\theta)^2$. Hence, $E \exp\{\theta(Z^2 - 1)\} \le \exp(2\theta^2)$. $\square$

Lemma 2. For a set of random variables X_i, $i = 1, \ldots, K$, and an event A, if there exists two positive constants C_1 and C_2 such that for all $|\theta| \le C_1$,

$$E(\exp(\theta X_i)I_A) \le \exp(C_2\theta^2), \qquad (A.2)$$

then for w_i's being weights satisfying $\sum_{i=1}^{K} |w_i| \le w \in [1, \infty)$, we have

$$P\left(\left\{ \left| \sum_{i=1}^{K} w_i X_i \right| > x \right\} \cap A \right)$$
$$\le 2\exp\left(-\frac{x^2}{4C_2 w^2} \right), \quad \text{when } 0 \le x \le 2C_1 C_2.$$

Proof. By the Markov inequality, for $0 \le \xi \le C_1/w$ and $w^* = \sum_{i=1}^{K} |w_i|$, we have

$$P\left(\left\{ \left| \sum_{i=1}^{K} w_i X_i \right| > x \right\} \cap A \right)$$
$$\le \exp(-\xi x)E\left(\exp\left(\xi \left| \sum_{i=1}^{K} w_i X_i \right| \right) I_A \right)$$
$$\le \exp(-\xi x)w^{*-1} \sum_{i=1}^{K} |w_i| E\left(\exp(\xi w |X_i|)I_A \right)$$
$$\le 2\exp\left(C_2 \xi^2 w^2 - \xi x \right). \qquad (A.3)$$

Taking $\xi = x/(2C_2 w^2)$, we have

$$P\left(\left\{ \left| \sum_{i=1}^{K} w_i X_i \right| > x \right\} \cap A \right) \le 2\exp\left(-\frac{x^2}{4C_2 w^2} \right),$$

when $0 \le x \le 2C_1 C_2$. $\qquad (A.4)$

$\square$

Lemma 3. (**A Concentration Inequality for Realized Volatility**) Let $[\widetilde{X, X}]_1 = \sum_{i=1}^{n}(X_{v_i} - X_{v_{i-1}})^2$ be the realized volatility based on the discretely observed X process from model (1) of the univariate case: $dX_t = \mu_t dt + \sigma_t dW_t$. Under Conditions 1–3,

(A) if $\sigma_t^{(X)} \le C_\sigma < \infty$ for all $t \in [0, 1]$, then for all large n, for $x \in [0, c\sqrt{n}]$,

$$P\left\{ n^{1/2}\left| [\widetilde{X, X}]_1 - \int_0^1 \sigma_t^2 dt \right| > x \right\} \le 2\exp\{-Cx^2\},$$

where the constants c and C can be taken as in (A.7).

(B) If the tail behavior of $\sigma_t^{(X)}$ satisfies,

$$P\left\{ \sup_{0 \le t \le 1} \sigma_t^{(X)} \ge C_\sigma \right\} \le k_\sigma \exp\left\{ -aC_\sigma^b \right\}, \quad \text{for any } C_\sigma > 0$$

with positive constants k_σ, a, and b, then for all large n, for $x \in [0, cn^{\frac{4+b}{2b}}]$,

$$P\left\{ n^{1/2}\left| [\widetilde{X, X}]_1 - \int_0^1 \sigma_t^2 dt \right| > x \right\}$$
$$\le (2 + k_\sigma)\exp\left\{ -Cx^{\frac{2b}{4+b}} \right\}.$$

A set of candidate values for c and C are given in (A.8).

Proof. For any constant $C_\sigma > 0$, define a stopping time $\Gamma_{C_\sigma} := \inf\{t : \sup_{0 \le s \le t} \sigma_s > C_\sigma\} \wedge 1$. Let

$$\tilde{\sigma}_s = \begin{cases} \sigma_s, & \text{when } s \le \Gamma_{C_\sigma} \\ C_\sigma, & \text{when } s > \Gamma_{C_\sigma} \end{cases}$$

and $\tilde{X}_t = \int_0^t \tilde{\sigma}_s dW_s$. By time change for martingales, [see, for example, Theorem 4.6 of chapter 3 of Karatzas and Shreve (2000)], if $\tau_t = \inf\{s : [\tilde{X}]_s \ge t\}$ where $[\tilde{X}]_s$ is the quadratic variation process, then $B_t := \tilde{X}_{\tau_t}$ is a Brownian-motion with regard to $\{\mathcal{F}_{\tau_t}\}_{0 \le t \le \infty}$. We then have that

$$E \exp\left(\theta\left(\tilde{X}_t^2 - \int_0^t \tilde{\sigma}_s^2 ds \right) \right) = E \exp\left(\theta\left(B_{[\tilde{X}]_t}^2 - [\tilde{X}]_t \right) \right).$$

Note further that for any t, $[\tilde{X}]_t$ is a stopping time with regard to $\{\mathcal{F}_{\tau_s}\}_{0 \le s \le \infty}$, and the process $\exp(\theta(B_s^2 - s))$ is a submartingale for any θ. By the optional sampling theorem, using $[\tilde{X}]_u \le C_\sigma^2 u$ (bounded stopping time), we have

$$E \exp\left(\theta\left(B_{[\tilde{X}]_u}^2 - [\tilde{X}]_u \right) \right) \le E \exp\left(\theta\left(B_{C_\sigma^2 u}^2 - C_\sigma^2 u \right) \right).$$

Therefore, note that $\Delta X_i = \Delta \tilde{X}_i$ on the set of $\{\Gamma_{C_\sigma} = 1\}$, we have that, under Condition 3,

$$E\left(\exp\left\{ \theta\sqrt{n}\left((\Delta X_i)^2 - \int_{v_{i-1}}^{v_i} \sigma_t^2 dt \right) \right\} I_{\{\Gamma_{C_\sigma} = 1\}} | \mathcal{F}_{v_{i-1}} \right)$$
$$\le E \exp\left\{ \theta\sqrt{n}\left(B_{\frac{C_\sigma^2 C_\Delta}{n}}^2 - \frac{C_\sigma^2 C_\Delta}{n} \right) \right\}$$
$$= E \exp\left\{ \theta\frac{C_\sigma^2 C_\Delta}{\sqrt{n}}(Z^2 - 1) \right\}, \qquad (A.5)$$

where $Z \sim N(0, 1)$ and $\Delta X_i = X_{v_i} - X_{v_{i-1}}$.

It follows from the law of iterated expectations and (A.5) that

$$
E\left(\exp\left\{\theta\sqrt{n}\left(\widetilde{[X, X]}_1 - \int_0^1 \sigma_t^2 dt\right)\right\} I_{\{\Gamma_{C_\sigma}=1\}}\right)
$$

$$
= E\left[\left(\exp\left\{\theta\sqrt{n}\left(\sum_{i=1}^{n-1}(\Delta X_i)^2 - \int_0^{v_{n-1}}\sigma_t^2 dt\right)\right\} I_{\{\Gamma_{C_\sigma}=1\}}\right)\right.
$$

$$
\left.\cdot E\left(\exp\left\{\theta\sqrt{n}\left(\Delta X_n^2 - \int_{v_{n-1}}^{v_n}\sigma_t^2 dt\right)\right\} I_{\{\Gamma_{C_\sigma}=1\}}|\mathcal{F}_{v_{n-1}}\right)\right]
$$

$$
\leq E\left(\exp\left\{\theta\sqrt{n}\left(\sum_{i=1}^{n-1}(\Delta X_i)^2 - \int_0^{v_{n-1}}\sigma_t^2 dt\right)\right\} I_{\{\Gamma_{C_\sigma}=1\}}\right)
$$

$$
\times E\exp\left\{\theta\frac{C_\sigma^2 C_\Delta}{\sqrt{n}}(Z^2-1)\right\},
$$

where $Z \sim N(0, 1)$. Repeating the process above, we obtain

$$
E\left(\exp\left\{\theta\sqrt{n}\left(\widetilde{[X, X]}_1 - \int_0^1 \sigma_t^2 dt\right)\right\} I_{\{\Gamma_{C_\sigma}=1\}}\right)
$$

$$
\leq \left(E\exp\left\{\theta\frac{C_\sigma^2 C_\Delta}{\sqrt{n}}(Z^2-1)\right\}\right)^n.
$$

By Lemma 1, we have for $|\theta| \leq \frac{\sqrt{n}}{4C_\sigma^2 C_\Delta}$,

$$
E\left(\exp\left\{\theta\sqrt{n}\left(\widetilde{[X, X]}_1 - \int_0^1 \sigma_t^2 dt\right)\right\} I_{\{\Gamma_{C_\sigma}=1\}}\right) \leq \exp\left\{2\theta^2 C_\sigma^4 C_\Delta^2\right\}.
\tag{A.6}
$$

By Lemma 2, we have,

$$
P\left(\left\{n^{1/2}\left|\widetilde{[X, X]}_1 - \int_0^1 \sigma_t^2 dt\right| > x\right\} \cap \{\Gamma_{C_\sigma}=1\}\right)
$$

$$
\leq 2\exp\left\{-\frac{x^2}{8C_\sigma^4 C_\Delta^2}\right\},
\tag{A.7}
$$

when $0 \leq x \leq C_\sigma^2 C_\Delta \sqrt{n}$. This proves the first half of the theorem. For the second half,

$$
P\left\{n^{1/2}\left|\widetilde{[X, X]}_1 - \int_0^1 \sigma_t^2 dt\right| > x\right\}
$$

$$
\leq P\left\{n^{1/2}\left|\widetilde{[X, X]}_1 - \int_0^1 \sigma_t^2 dt\right| > x, \Gamma_{C_\sigma}=1\right\} + P\{\Gamma_{C_\sigma}<1\}
$$

$$
\leq \exp\{-\theta x\} E\left(\exp\left\{\theta\sqrt{n}\left(\left|\widetilde{[X, X]}_1 - \int_0^1 \sigma_t^2 dt\right|\right)\right\} I_{\{\Gamma_{C_\sigma}=1\}}\right)
$$

$$
+ P\{\Gamma_{C_\sigma}<1\} \quad \text{for nonnegative } \theta
$$

$$
\leq 2\exp\left\{-\frac{x^2}{8C_\sigma^4 C_\Delta^2}\right\} + k_\sigma \exp\left\{-aC_\sigma^b\right\},
$$

$$
\text{when } 0 \leq x \leq C_\sigma^2 C_\Delta \sqrt{n}.
$$

Now, let $C_\sigma = (\frac{x^2}{8aC_\Delta^2})^{\frac{1}{4+b}}$, we have that when $0 \leq x \leq (8a)^{\frac{-2}{b}}\cdot C_\Delta \cdot n^{\frac{4+b}{2b}}$,

$$
P\left\{n^{1/2}\left|\widetilde{[X, X]}_1 - \int_0^1 \sigma_t^2 dt\right| > x\right\}
$$

$$
\leq (2+k_\sigma)\exp\left\{-8^{\frac{-b}{4+b}}\cdot a^{\frac{4}{4+b}}\cdot C_\Delta^{\frac{-2b}{4+b}}\cdot x^{\frac{2b}{4+b}}\right\}.
\tag{A.8}
$$

$\square$

Proof of Theorem 1

We first prove the results conditioning on the set of observation times $\mathcal{V}$. Recall notation introduced in Sections 3.2 and 3.3. Let n be the observation frequency. For simplicity of notation, without ambiguity, we will write $\tau_{n,i}$ as τ_i and $\sigma_t^{(X)}$ as σ_t. Again let $\Gamma_{C_\sigma} := \inf\{t : \sup_{0 \leq s \leq t}\sigma_s > C_\sigma\} \wedge 1$. Denote the TSRV based on the latent process by

$$
\widetilde{\langle X, X\rangle}_1^{(K)} = \widetilde{[X, X]}_1^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}\widetilde{[X, X]}_1^{(J)},
\tag{A.9}
$$

where $\widetilde{[X, X]}_1^{(K)} = K^{-1}\sum_{i=K}^n(X_{\tau_i} - X_{\tau_{i-K}})^2$. Then, from the definition, we have,

$$
\widetilde{\langle X, X\rangle}_1 = \widetilde{[X, X]}_1^{(K)} + \widetilde{[\epsilon^X, \epsilon^X]}_1^{(K)} + 2\widetilde{[X, \epsilon^X]}_1^{(K)}
$$

$$
- \frac{\bar{n}_K}{\bar{n}_J}\left(\widetilde{[X, X]}_1^{(J)} + \widetilde{[\epsilon^X, \epsilon^X]}_1^{(J)} + 2\widetilde{[X, \epsilon^X]}_1^{(J)}\right)
$$

$$
= \frac{1}{K}\sum_{l=0}^{K-1}V_K^{(l)} - \frac{\bar{n}_K}{\bar{n}_J}\widetilde{[X, X]}_1^{(J)} + R_1 + R_2,
\tag{A.10}
$$

where $R_1 = \widetilde{[\epsilon^X, \epsilon^X]}_1^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}\widetilde{[\epsilon^X, \epsilon^X]}_1^{(1)}$, $R_2 = 2\widetilde{[X, \epsilon^X]}_1^{(K)} - 2\frac{\bar{n}_K}{\bar{n}_J}\widetilde{[X, \epsilon^X]}_1^{(1)}$, and

$$
V_K^{(l)} = \sum_{j=1}^{\bar{n}_K}(X_{\tau_{jK+l}} - X_{\tau_{(j-1)K+l}})^2, \quad \text{for } l = 0, 1, \ldots, K-1.
$$

Note that we have assumed that $\bar{n}_K = \frac{n-K+1}{K}$ is an integer above, to simplify the presentation.

Recall that we consider the case when $J = 1$, or $\bar{n}_J = n$. Let

$$
I_1 = \frac{1}{K}\sum_{l=0}^{K-1}\sqrt{\bar{n}_K}\left(V_K^{(l)} - \int_0^1\sigma_t^2 dt\right) - \left(\frac{\bar{n}_K}{n}\right)^{\frac{3}{2}}\cdot\sqrt{n}\left(\widetilde{[X, X]}_1^{(1)}\right.
$$

$$
\left. - \int_0^1\sigma_t^2 dt\right) + \sqrt{\bar{n}_K}R_1 + \sqrt{\bar{n}_K}R_2,
\tag{A.11}
$$

and $I_2 = -\frac{\bar{n}_K^{3/2}}{n}\int_0^1\sigma_t^2 dt$. We are interested in

$$
\sqrt{\bar{n}_K}\left(\widetilde{\langle X, X\rangle}_1 - \int_0^1\sigma_t^2 dt\right) = I_1 + I_2.
$$

The key idea is to compute the moment generating functions for each term in I_1 and then to use Lemma 2 to conclude.

For the first term in I_1, since $V_k^{(l)}$ is a realized volatility based on discretely observed X process, with observation frequency satisfying $\sup_{1 \leq i \leq \bar{n}_K}\bar{n}_K\cdot(\tau_{iK+l} - \tau_{(i-1)K+l}) \leq C_\Delta$, we have, by (A.6) in Lemma 3, for $|\theta| \leq \frac{\sqrt{\bar{n}_K}}{4C_\sigma^2 C_\Delta}$,

$$
E\left(\exp\left\{\theta\sqrt{\bar{n}_K}(V_K^{(l)} - \int_0^1\sigma_t^2 dt)\right\} I_{\{\Gamma_{C_\sigma}=1\}}\right) \leq \exp\left\{2\theta^2 C_\sigma^4 C_\Delta^2\right\}.
\tag{A.12}
$$

For the second term in I_1, we have obtained in (A6) that

$$
E\left(\exp\left\{\theta\sqrt{n}\left(\widetilde{[X, X]}_1^{(1)} - \int_0^1\sigma_t^2 dt\right)\right\} I_{\{\Gamma_{C_\sigma}=1\}}\right)
$$

$$
\leq \exp\left\{2\theta^2 C_\sigma^4 C_\Delta^2\right\}, \quad \text{when } |\theta| \leq \frac{\sqrt{n}}{4C_\sigma^2 C_\Delta}.
\tag{A.13}
$$

We introduce an auxiliary sequence a_n that grows with n at a moderate rate to facilitate our presentation in the following. In particular, we can set $a_n = n^{1/12}$. Let us now deal with R_1, the third term in I_1. Note that from the definition

$$
\begin{aligned}
&\sqrt{\bar{n}_K} R_1 \\
&= \frac{\sqrt{\bar{n}_K}}{K} \left\{ \sum_{i=K}^{n} (\epsilon_i - \epsilon_{i-K})^2 - \frac{n-K+1}{n} \sum_{i=1}^{n} (\epsilon_i - \epsilon_{i-1})^2 \right\} \\
&= \frac{\sqrt{\bar{n}_K}(n-K+1)}{K\sqrt{n}} \cdot \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i \epsilon_{i-1} - \frac{\sqrt{\bar{n}_K}\sqrt{n-K+1}}{K} \\
&\quad \cdot \frac{2}{\sqrt{n-K+1}} \sum_{i=K}^{n} \epsilon_i \epsilon_{i-K} - \frac{\sqrt{\bar{n}_K}\sqrt{K-1}a_n}{K} \\
&\quad \cdot \frac{1}{a_n\sqrt{K-1}} \sum_{i=1}^{K-1} \left(\epsilon_i^2 - \eta_X^2 \right) - \frac{\sqrt{\bar{n}_K}\sqrt{K-1}a_n}{K} \\
&\quad \cdot \frac{1}{a_n\sqrt{K-1}} \sum_{n-K+1}^{n-1} \left(\epsilon_i^2 - \eta_X^2 \right) + \frac{\sqrt{\bar{n}_K}(K-1)a_n}{K\sqrt{n}} \\
&\quad \cdot \frac{1}{a_n\sqrt{n}} \sum_{i=1}^{n} \left(\epsilon_i^2 - \eta_X^2 \right) + \frac{\sqrt{\bar{n}_K}(K-1)a_n}{K\sqrt{n}} \cdot \frac{1}{a_n\sqrt{n}} \sum_{i=0}^{n-1} \left(\epsilon_i^2 - \eta_X^2 \right).
\end{aligned}
$$
(A.14)

The first two terms in (A.14) are not sums of independent variables. But they can be decomposed into sums of independent random variables and the moment generating functions can be computed. To simplify the argument without losing the essential ingredient, let us focus on the first term of (A.14). We have the following decomposition

$$
\sum_{i=1}^{n} \varepsilon_i \varepsilon_{i-1} = \sum_{\text{odd }i} \varepsilon_i \varepsilon_{i-1} + \sum_{\text{even }i} \varepsilon_i \varepsilon_{i-1},
$$

and the summands in each terms of the right-hand side are now independent. Therefore, we need only to calculate the moment generating function of $\varepsilon_i \varepsilon_{i-1}$.

For two independent normally distributed random variables $X \sim N(0, \sigma_X^2)$ and $Y \sim N(0, \sigma_Y^2)$, it can easily be computed that

$$
\begin{aligned}
E(\exp\{\theta n^{-1/2} XY\}) &= \left(\frac{1}{1 - \sigma_X^2 \sigma_Y^2 \theta^2/n} \right)^{1/2} \\
&\leq \exp \left\{ \sigma_X^2 \sigma_Y^2 \theta^2/n \right\} \quad \text{when } |\theta| \leq \frac{\sqrt{n}}{\sqrt{2}\sigma_X\sigma_Y},
\end{aligned}
$$

where we have used the fact that $\log(1-x) \geq -2x$ when $0 \leq x \leq \frac{1}{2}$.

Hence, by independence, it follows that (we assume n is even to simplify the presentation)

$$
\begin{aligned}
E \exp \left\{ 2\theta n^{-1/2} \sum_{\text{odd }i} \varepsilon_i \varepsilon_{i-1} \right\} &= \left(\frac{1}{1 - 4\eta_X^4 \theta^2/n} \right)^{n/4} \\
&\leq \exp \left\{ 2\eta_X^4 \theta^2 \right\}, \quad \text{when } |\theta| \leq \frac{\sqrt{n}}{2\sqrt{2}\eta_X^2}.
\end{aligned}
$$
(A.15)

The second term in R_1 works similarly and has the same bound. For example, when $\bar{n}_K$ is even, one can have the following decomposition

$$
\sum_{i=K}^{n} \epsilon_i \epsilon_{i-K} = \sum_{j=1}^{\bar{n}_K/2} \sum_{i=2jK-K}^{2jK-1} \epsilon_i \epsilon_{i-K} + \sum_{j=1}^{\bar{n}_K/2} \sum_{i=2jK}^{2jK+K-1} \epsilon_i \epsilon_{i-K}.
$$

The last four terms are sums of independent χ^2-distributed random variables and their moment generating functions can easily be bounded

by using Lemma 1. Taking the term $\frac{1}{a_n\sqrt{K-1}} \sum_{i=1}^{K-1} (\epsilon_i^2 - \eta_X^2)$ for example, we have

$$
\begin{aligned}
&E \left(\exp \left\{ \frac{\theta}{a_n\sqrt{K-1}} \sum_{i=1}^{K-1} \left(\epsilon_i^2 - \eta_X^2 \right) \right\} \right) \\
&\qquad \leq \exp \left\{ 2\eta_X^4 \theta^2/a_n^2 \right\} \quad \text{when } |\theta| \leq \frac{a_n\sqrt{K-1}}{4\eta_X^2}.
\end{aligned}
$$

For the term R_2, we have,

$$
\begin{aligned}
\sqrt{\bar{n}_K} R_2 &= \frac{2a_n\bar{n}_K}{n} \frac{1}{a_n} \left(\sum_{i=1}^{n} \Delta X_i \epsilon_{i-1} - \sum_{i=1}^{n} \Delta X_i \epsilon_i \right) \\
&\quad + \frac{2}{a_n} \frac{a_n\sqrt{\bar{n}_K}}{K} \left(\sum_{i=K}^{n} \Delta^{(K)} X_i \epsilon_i - \sum_{i=K}^{n} \Delta^{(K)} X_i \epsilon_{i-K} \right),
\end{aligned}
$$
(A.16)

where $\Delta X_i = X_{\tau_i} - X_{\tau_{i-1}}$, and $\Delta^{(K)} X_i = X_{\tau_i} - X_{\tau_{i-K}}$. The first term above satisfies

$$
\begin{aligned}
&E \left(\exp \left\{ \frac{\theta}{a_n} \sum_{i=1}^{n} \Delta X_i \epsilon_i \right\} I_{\{\Gamma_{C_\sigma}=1\}} \right) \\
&= E \left(\exp \left\{ \sum_{i=1}^{n} \left(\frac{\theta}{a_n} \Delta X_i \right)^2 \eta_X^2 / 2 \right\} I_{\{\Gamma_{C_\sigma}=1\}} \right) \\
&\leq \left(E \left(\exp \left\{ \theta^2 \eta_X^2 C_\sigma^2 C_\Delta Z^2 / 2na_n^2 \right\} \right) \right)^n \\
&= \left(\frac{1}{1 - \eta_X^2 C_\sigma^2 C_\Delta \theta^2/na_n^2} \right)^{n/2} \\
&\leq \exp \left\{ \eta_X^2 C_\sigma^2 C_\Delta \theta^2 / a_n^2 \right\}, \quad \text{when } |\theta| \leq \frac{\sqrt{n}a_n}{\sqrt{2C_\Delta C_\sigma}\eta_X}, \quad (\text{A}.17)
\end{aligned}
$$

where in the second line we have again used the optional sampling theorem and law of iterated expectations as in the derivations of Lemma 3; Z denotes a standard normal random variable. The second term in R_2 works similarly and has the same bound. For the third term, by conditioning on the X-process first, we have

$$
\begin{aligned}
&E \left[\exp \left\{ \frac{a_n\theta\sqrt{\bar{n}_K}}{K} \sum_{i=K}^{n} \Delta^{(K)} X_i \epsilon_i \right\} I_{\{\Gamma_{C_\sigma}=1\}} \right] \\
&= E \left[\exp \left\{ \frac{a_n^2\theta^2\bar{n}_K}{2K^2} \sum_{l=0}^{K-1} \sum_{j=1}^{\bar{n}_K} \left(\Delta^{(K)} X_{jK+l} \right)^2 \eta_X^2 \right\} I_{\{\Gamma_{C_\sigma}=1\}} \right] \\
&\leq \Pi_{l=0}^{K-1} \left\{ E \left[\exp \left\{ \frac{a_n^2\theta^2\bar{n}_K\eta_X^2}{2K} \sum_{j=1}^{\bar{n}_K} \left(\Delta^{(K)} X_{jK+l} \right)^2 \right\} \right. \right. \\
&\quad \left. \left. \times I_{\{\Gamma_{C_\sigma}=1\}} \right] \right\}^{\frac{1}{K}} \leq \Pi_{l=0}^{K-1} \left\{ \left(1 - \frac{a_n^2\theta^2\eta_X^2}{K} C_\sigma^2 C_\Delta \right)^{-\bar{n}_K/2} \right\}^{\frac{1}{K}} \\
&\leq \exp \left\{ \frac{a_n^2\theta^2\bar{n}_K\eta_X^2}{K} C_\sigma^2 C_\Delta \right\} \quad \text{when } |\theta| \leq \frac{\sqrt{K}}{\sqrt{2C_\Delta}a_n\eta C_\sigma},
\end{aligned}
$$
(A.18)

where we have used the Hölder's inequality above. The fourth term works similarly and has the same bound.

Combining the results for all the terms (A.12)–(A.18) together, applying Lemma 2 to I_1, we conclude that the conditions for Lemma 2 are satisfied with $A = \{\Gamma_{C_\sigma} = 1\}$, $C_1 = C_{1,x}\sqrt{\bar{n}_K}$,

$$
C_{1,x} = \min \left\{ \frac{1}{4C_\sigma^2 C_\Delta}, \frac{\sqrt{n/\bar{n}_K}}{2\sqrt{2}\eta_X^2}, \frac{a_n\sqrt{(K-1)/\bar{n}_K}}{4\eta_X^2}, \right.
$$

$$\left.\frac{a_n\sqrt{n/\bar{n}_K}}{\sqrt{2C_\Delta}\eta_X C_\sigma}, \frac{\sqrt{K/\bar{n}_K}}{\sqrt{2C_\Delta}a_n\eta C_\sigma}\right\} \tag{A.19}$$

$$= \frac{1}{4C_\sigma^2 C_\Delta} \text{ for big enough } n$$

and

$$C_2 = \max\left\{2C_\sigma^4 C_\Delta^2, 2\eta_X^4, 2\eta_X^4/a_n^2, \eta_X^2 C_\sigma^2 C_\Delta/a_n^2,\right.$$
$$\left.\frac{a_n^2\bar{n}_K\eta_X^2}{K}C_\sigma^2 C_\Delta\right\} \tag{A.20}$$

$$= \max\left\{2C_\sigma^4 C_\Delta^2, 2\eta_X^4\right\} \text{ for big enough } n$$

$$= 2C_\sigma^4 C_\Delta^2 \text{ for the typical case when } C_\Delta \geq 1 \text{ and } C_\sigma \geq \eta_X.$$

Let $w = 8$, which is larger, when n is sufficiently large, than

$$\underbrace{\frac{1}{K}\sum_{l=0}^{K-1}1 + \left(\frac{\bar{n}_K}{n}\right)^{3/2}}_{\text{coef. in the first two terms of } I_1}$$

$$+ \underbrace{\frac{2\sqrt{\bar{n}_K}\sqrt{n}}{K} + \frac{2\sqrt{\bar{n}_K}a_n}{\sqrt{K}} + \frac{2\sqrt{\bar{n}_K}a_n}{\sqrt{n}}}_{\text{controls coefficients in (A.14)}}$$

$$+ \underbrace{\frac{4a_n\bar{n}_K}{n} + \frac{4}{a_n}}_{\text{coefficients in (A.16)}}.$$

Set $C_{I_1} = (4C_2 w^2)^{-1}$. By Lemma 2, when $0 \leq x \leq 2C_{1,x}C_2\sqrt{\bar{n}_K}$,

$$P\left(\{|I_1| > x\} \cap \{\Gamma_{C_\sigma} = 1\}\right) \leq 2\exp(-C_{I_1}x^2).$$

Hence when $0 \leq x \leq 4C_{1,x}C_2\sqrt{\bar{n}_K}$,

$$P\left(\left\{|I_1| > \frac{x}{2}\right\} \cap \{\Gamma_{C_\sigma} = 1\}\right) \leq 2\exp\left(-\frac{C_{I_1}}{4}x^2\right). \tag{A.21}$$

For the term I_2, let $C_{I_2} = 2^{3/2}C_\sigma^2$, we have, by Condition 4,

$$I_2 = \frac{\bar{n}_K^{3/2}}{n}\int_0^1 \sigma_t^2 dt \leq C_{I_2}/\sqrt{n}, \text{ on the set } \{\Gamma_{C_\sigma} = 1\}. \tag{A.22}$$

Hence

$$P\left(\left\{|I_2| > \frac{x}{2}\right\} \cap \{\Gamma_{C_\sigma} = 1\}\right) \leq \begin{cases} 0, & \text{if } x > 2C_{I_2}/\sqrt{n} \\ 1, & \text{if } x \leq 2C_{I_2}/\sqrt{n}. \end{cases}$$

Since for all large n, $1 \leq 2\exp(-\frac{C_{I_1}4C_{I_2}^2}{4n})$, which is smaller than $2\exp(-\frac{C_{I_1}x^2}{4})$ when $x \leq \frac{2C_{I_2}}{\sqrt{n}}$, we have that for all large n,

$$P\left(\left\{|I_2| > \frac{x}{2}\right\} \cap \{\Gamma_{C_\sigma} = 1\}\right) \leq 2\exp\left(-\frac{C_{I_1}}{4}x^2\right). \tag{A.23}$$

Combining (A.21) and (A.23), we have, when $0 \leq x \leq 4C_{1,x}C_2\sqrt{\bar{n}_K}$,

$$P\left(\left\{\sqrt{\bar{n}_K}\left|\widehat{\langle X, X\rangle}_1 - \int_0^1\sigma_t^2 dt\right| > x\right\} \cap \{\Gamma_{C_\sigma} = 1\}\right)$$

$$= P\left(\{|I_1 + I_2| > x\} \cap \{\Gamma_{C_\sigma} = 1\}\right)$$

$$\leq P\left(\{|I_1| > x/2\} \cap \{\Gamma_{C_\sigma} = 1\}\right)$$

$$+ P\left(\{|I_2| > x/2\} \cap \{\Gamma_{C_\sigma} = 1\}\right) \leq 4\exp\left(-\frac{C_{I_1}}{4}x^2\right).$$

By Condition 4 again, we have, when $0 \leq x \leq cn^{1/6}$,

$$P\left(\left\{n^{1/6}\left|\widehat{\langle X, X\rangle}_1 - \int_0^1\sigma_t^2 dt\right| > x\right\} \cap \{\Gamma_{C_\sigma} = 1\}\right)$$

$$\leq P\left(\left\{\sqrt{\bar{n}_K}\left|\widehat{\langle X, X\rangle}_1 - \int_0^1\sigma_t^2 dt\right| > x/\sqrt{2}\right\} \cap \{\Gamma_{C_\sigma} = 1\}\right)$$

$$\leq 4\exp(-Cx^2), \tag{A.24}$$

where $c = 4\sqrt{2}C_{1,x}C_2\sqrt{\bar{n}_K}$, and $C = \frac{C_{I_1}}{8} = (32C_2 w^2)^{-1}$. For big enough n and the typical case when $C_\Delta \geq 1$ and $C_\sigma \geq \eta_X$, we have

$$c = 2\sqrt{2}C_\sigma^2 C_\Delta \quad \text{and} \quad C = \frac{1}{64w^2 C_\sigma^4 C_\Delta^2}. \tag{A.25}$$

Notice that this conditional result depends only on the observation frequency n and not on the locations of the observation times as long as Condition 3 is satisfied, (A.24) holds unconditionally on the set of the observation times. This proves the first half of the Theorem 1 when $\Gamma_{C_\sigma} \equiv 1$.

For the second half of the theorem, we have

$$P\left(n^{1/6}\left|\widehat{\langle X, X\rangle}_1 - \int_0^1\sigma_t^2 dt\right| > x\right)$$

$$= P\left(\left\{n^{1/6}\left|\widehat{\langle X, X\rangle}_1 - \int_0^1\sigma_t^2 dt\right| > x\right\} \cap \{\Gamma_{C_\sigma} = 1\}\right)$$

$$+ P\left(\{\Gamma_{C_\sigma} < 1\}\right)$$

$$< 4\exp\left(-\frac{x^2}{64w^2 C_\sigma^4 C_\Delta^2}\right) + k_\sigma\exp\left\{-aC_\sigma^b\right\},$$

$$\text{when } 0 \leq x \leq 2\sqrt{2}C_\sigma^2 C_\Delta n^{1/6}.$$

Letting $C_\sigma = \left(\frac{x^2}{64w^2 a C_\Delta^2}\right)^{\frac{1}{4+b}}$, we have, when $0 \leq x \leq 2^{\frac{3b-12}{2b}}C_\Delta(w^2 a)^{\frac{-2}{b}}\cdot n^{\frac{4+b}{6b}}$,

$$P\left\{n^{1/6}\left|\widetilde{[X, X]}_1 - \int_0^1\sigma_t^2 dt\right| > x\right\}$$

$$\leq (4 + k_\sigma)\exp\left\{-(64w^2)^{\frac{-b}{4+b}}\cdot a^{\frac{4}{4+b}}\cdot C_\Delta^{\frac{-2b}{4+b}}\cdot x^{\frac{2b}{4+b}}\right\}. \tag{A.26}$$

Remark 3. In the above proof, we have demonstrated by using a sequence a_n that goes to ∞ at a moderate rate that one can eliminate the impact of the small order terms on the choices of the constants, as long as the terms have their moment generating functions satisfy inequalities of form (A.2). We will use this technique again in the next subsection.

Proof of Theorem 2

We again conduct all the analysis assuming that the observation times are given. Our final result holds because the conditional result does not depend on the locations of the observation times as long as Condition 3 is satisfied.

Recall notation for the observation times as introduced in Section 3.2. Define

$$Z^+ = X + Y \quad \text{and} \quad Z^- = X - Y.$$

Z^+ and Z^- are diffusion processes with volatilities satisfying Condition 2. To see this, let W^+ and W^- be processes such that

$$dW_t^+ = \frac{\sigma_t^{(X)}dB_t^{(X)} + \sigma_t^{(Y)}dB_t^{(Y)}}{\sqrt{\left(\sigma_t^{(X)}\right)^2 + \left(\sigma_t^{(Y)}\right)^2 + 2\rho_t\sigma_t^{(X)}\sigma_t^{(Y)}}}$$

and

$$dW_t^- = \frac{\sigma_t^{(X)}dB_t^{(X)} - \sigma_t^{(Y)}dB_t^{(Y)}}{\sqrt{\left(\sigma_t^{(X)}\right)^2 + \left(\sigma_t^{(Y)}\right)^2 - 2\rho_t\sigma_t^{(X)}\sigma_t^{(Y)}}}.$$

W^+ and W^- are standard Brownian motions by Levy's characterization of Brownian motion [see, for example, Theorem 3.16, chapter 3 of Karatzas and Shreve (2000)]. Write

$$\sigma_t^{Z^+} = \sqrt{\left(\sigma_t^{(X)}\right)^2 + \left(\sigma_t^{(Y)}\right)^2 + 2\rho_t\sigma_t^{(X)}\sigma_t^{(Y)}}$$

and

$$\sigma_t^{Z^-} = \sqrt{\left(\sigma_t^{(X)}\right)^2 + \left(\sigma_t^{(Y)}\right)^2 - 2\rho_t\sigma_t^{(X)}\sigma_t^{(Y)}},$$

we have

$$dZ^+ = \sigma_t^{Z^+}dW_t^+ \quad \text{and} \quad dZ^- = \sigma_t^{Z^-}dW_t^-$$

with

$$0 \leq \sigma_t^{Z^+}, \sigma_t^{Z^-} \leq 2C_\sigma$$

when $\sigma_t^{(X)}, \sigma_t^{(Y)}$ are bounded by C_σ, or

$$P\left\{\sup_{0\leq t\leq 1}\sigma_t^{Z^+} \geq 2C_\sigma\right\} \leq P\left\{\sup_{0\leq t\leq 1}\sigma_t^X \geq C_\sigma\right\}$$
$$+ P\left\{\sup_{0\leq t\leq 1}\sigma_t^Y \geq C_\sigma\right\} \leq 2k_\sigma \exp\left\{-aC_\sigma^b\right\},$$

when σ_t^X, σ_t^Y are such that their tail probabilities are bounded as in Condition 2.

In fact, let $\Gamma_{C_\sigma} := \inf\{t : \sup_{0\leq s\leq t}\sigma_s^{(X)} > C_\sigma \text{ or } \sup_{0\leq s\leq t}\sigma_s^{(Y)} > C_\sigma\} \wedge 1$. We have $P\{\Gamma_{C_\sigma} < 1\} \leq P\{\sup_{0\leq t\leq 1}\sigma_t^X \geq C_\sigma\} + P\{\sup_{0\leq t\leq 1}\sigma_t^Y \geq C_\sigma\} \leq 2k_\sigma \exp\{-aC_\sigma^b\}$.

For the observed Z^+ and Z^- processes, we have

$$Z_{v_i}^{+,o} = X_{t_i}^o + Y_{s_i}^o = Z_{v_i}^+ + \epsilon_{i,+} \text{ and}$$
$$Z_{v_i}^{-,o} = X_{t_i}^o - Y_{s_i}^o = Z_{v_i}^- + \epsilon_{i,-},$$

where t_i and s_i are the last ticks at or before v_i and

$$\epsilon_{i,+} = X_{t_i} - X_{v_i} + Y_{s_i} - Y_{v_i} + \epsilon_i^X + \epsilon_i^Y,$$
$$\epsilon_{i,-} = X_{t_i} - X_{v_i} - Y_{s_i} + Y_{v_i} + \epsilon_i^X - \epsilon_i^Y.$$

Note that $\widehat{\langle X, Y\rangle}_1 = \frac{1}{4}(\langle\widehat{Z^+, Z^+}\rangle_1 - \langle\widehat{Z^-, Z^-}\rangle_1)$. We can first prove analogous results as in Theorem 1 for $\langle\widehat{Z^+, Z^+}\rangle_1$ and $\langle\widehat{Z^-, Z^-}\rangle_1$, then use the results to obtain the final conclusion for TSCV.

For $\langle\widehat{Z^+, Z^+}\rangle_1$, the derivation is different from that of Theorem 1 only for the terms that involve the noise, namely $\sqrt{\bar{n}_K}R_1$ and $\sqrt{\bar{n}_K}R_2$. Write $\tilde{\Delta}X_i = X_{t_i} - X_{v_i}$ and $\tilde{\Delta}Y_i = Y_{s_i} - Y_{v_i}$. Then, the first term in $\sqrt{\bar{n}_K}R_1$ becomes

$$\frac{\sqrt{\bar{n}_K}\sqrt{\bar{n}}}{K}\cdot\frac{2}{\sqrt{\bar{n}}}\sum_{i=1}^{\bar{n}}\epsilon_{i,+}\epsilon_{i-1,+} = \frac{\sqrt{\bar{n}_K}\sqrt{\bar{n}}}{K}\cdot\frac{2}{\sqrt{\bar{n}}}$$

$$\times\sum_{i=1}^{\bar{n}}\left(\tilde{\Delta}X_i\tilde{\Delta}X_{i-1} + \tilde{\Delta}X_i\tilde{\Delta}Y_{i-1} + \tilde{\Delta}X_i\left(\epsilon_{i-1}^X + \epsilon_{i-1}^Y\right)\right.$$

$$+ \tilde{\Delta}Y_i\tilde{\Delta}X_{i-1} + \tilde{\Delta}Y_i\tilde{\Delta}Y_{i-1} + \tilde{\Delta}Y_i\left(\epsilon_{i-1}^X + \epsilon_{i-1}^Y\right) + \left(\epsilon_i^X + \epsilon_i^Y\right)$$

$$\times\tilde{\Delta}X_{i-1} + \left(\epsilon_i^X + \epsilon_i^Y\right)\tilde{\Delta}Y_{i-1} + \left(\epsilon_i^X + \epsilon_i^Y\right)\left(\epsilon_{i-1}^X + \epsilon_{i-1}^Y\right)\bigg).$$

The only $O_P(1)$ term is the last term, which involves only independent normals, and can be dealt with the same way as before (again assume $\bar{n}$ is even for simplicity of presentation):

$$E\exp\left\{2\theta\bar{n}^{-1/2}\sum_{\text{odd }i}\left(\epsilon_i^X + \epsilon_i^Y\right)\left(\epsilon_{i-1}^X + \epsilon_{i-1}^Y\right)\right\}$$

$$= E\exp\left\{2\theta\bar{n}^{-1/2}\sum_{\text{even }i}\left(\epsilon_i^X + \epsilon_i^Y\right)\left(\epsilon_{i-1}^X + \epsilon_{i-1}^Y\right)\right\}$$

$$= \left(\frac{1}{1 - 4\left(\eta_X^2 + \eta_Y^2\right)^2\theta^2/\bar{n}}\right)^{\bar{n}/4}$$

$$\leq \exp\left\{2\left(\eta_X^2 + \eta_Y^2\right)^2\theta^2\right\}, \text{ when } |\theta| \leq \frac{\sqrt{\bar{n}}}{2\sqrt{2}\left(\eta_X^2 + \eta_Y^2\right)}.$$

The other terms are of a smaller order of magnitude. By applying an $a_{\bar{n}}$ sequence which grows moderately with $\bar{n}$ as in Proof of Theorem 1 (we can set $a_{\bar{n}} = \bar{n}^{1/12}$), we can see easily that as long as we can show that the moment generating functions of these terms can indeed be suitably bounded as in (A.2), their exact bounds do not have effect on our choice of C_1, C_2, or ω. To show the bounds for the moment generating functions, first note that, for any positive number a and real valued b, by the optional sampling theorem (applied to submartingales $\exp(aB_s^2)$ and $\exp(b\tilde{\Delta}yB_s)$ with stopping time $[X]_u \leq C_\sigma^2 u$ for real number $\tilde{\Delta}y$), we have

$$E\left(\exp\{a(\tilde{\Delta}X_i)^2\}I_{\{\Gamma_{C_\sigma}=1\}}|\mathcal{F}_{i-1}\right)$$

$$\leq \left(E\left(\exp\left\{aC_\sigma^2C_\Delta Z^2/\bar{n}\right\}\right)\right) \text{ for } Z \sim N(0, 1)$$

$$= \left(\frac{1}{1 - 2aC_\sigma^2C_\Delta/\bar{n}}\right)^{1/2}, \tag{A.27}$$

where $\mathcal{F}_i$ is the information collected up to time v_i. Inequality (A.27) holds when $\tilde{\Delta}X_i$ is replaced by $\tilde{\Delta}Y_i$. Similarly,

$$E\left(\exp\{b\tilde{\Delta}X_i\tilde{\Delta}Y_{i-1}\}I_{\{\Gamma_{C_\sigma}=1\}}|\mathcal{F}_{i-2}\right)$$

$$\leq E\left(E(\exp\{b\tilde{\Delta}X_i\tilde{\Delta}Y_{i-1}\}I_{\{\Gamma_{C_\sigma}=1\}}|\mathcal{F}_{i-1})|\mathcal{F}_{i-2}\right)$$

$$\leq E\left(\exp\{b^2C_\Delta C_\sigma^2(\tilde{\Delta}Y_{i-1})^2/2\bar{n}\}I_{\{\Gamma_{C_\sigma}=1\}}|\mathcal{F}_{i-2}\right)$$

$$\leq \left(\frac{1}{1 - b^2C_\sigma^4C_\Delta^2/\bar{n}^2}\right)^{1/2}. \tag{A.28}$$

Inequalities (A.27) and (A.28) can be used to obtain the bounds we need. For example, by (A.28) and the law of iterated expectations,

$$E\left(\exp\left\{\theta\sum_{\text{odd}i}\tilde{\Delta}X_i\tilde{\Delta}Y_{i-1}\right\}I_{\{\Gamma_{C_\sigma}=1\}}\right) \leq \left(\frac{1}{1 - \theta^2C_\sigma^4C_\Delta^2/\bar{n}^2}\right)^{\bar{n}/4}$$

$$\leq \exp\left\{\theta^2C_\sigma^4C_\Delta^2\big/2\bar{n}\right\} \text{ when } |\theta| \leq \frac{\bar{n}}{\sqrt{2}C_\sigma^2C_\Delta},$$

and by independence, normality of the noise, the law of iterated expectations, and (A.27), we have

$$E\left(\exp\left\{\frac{\theta}{a_{\bar{n}}}\sum_{i=1}^{\bar{n}}\tilde{\Delta}X_i\left(\epsilon_{i-1}^X + \epsilon_{i-1}^Y\right)\right\}I_{\{\Gamma_{C_\sigma}=1\}}\right)$$

$$= E\left(\exp\left\{\sum_{i=1}^{\bar{n}}\left(\frac{\theta}{a_{\bar{n}}}\tilde{\Delta}X_i\right)^2\left(\eta_X^2 + \eta_Y^2\right)\big/2\right\}I_{\{\Gamma_{C_\sigma}=1\}}\right)$$

$$\leq \left(\frac{1}{1 - \left(\eta_X^2 + \eta_Y^2\right)\theta^2C_\sigma^2C_\Delta/\bar{n}a_{\bar{n}}^2}\right)^{\bar{n}/2}$$

$$\leq \exp\left\{\left(\eta_X^2 + \eta_Y^2\right)C_\sigma^2C_\Delta\theta^2/a_{\bar{n}}^2\right\},$$

when $|\theta| \le \dfrac{\sqrt{n}\,a_{\tilde{n}}}{C_\sigma\sqrt{2C_\Delta(\eta_X^2+\eta_Y^2)}}$.

Similar results can be found for the other terms above, with the same techniques.

The second term in $\sqrt{n_K}R_1$ works similarly and has the same bound. The other terms in $\sqrt{n_K}R_1$ (with η_X^2 replaced by $\eta_X^2+\eta_Y^2$) and the whole term of $\sqrt{n_K}R_2$ are of order $o_p(1)$ and have good tail behaviors. Again, by using a sequence $a_{\tilde{n}}$, we can conclude that their exact bounds will not matter in our choice of the constants and we only need to show that their moment generating functions are appropriately bounded as in (A.2). The arguments needed to prove the inequalities of form (A.2) for each elements in these terms are similar to those presented in the above proofs, and are omitted here.

Hence, by still letting $w=8$ and redefining

$$C_{1,x}=\frac{1}{4(2C_\sigma)^2C_\Delta} \quad\text{and}$$

$$C_2=\max\left\{2(2C_\sigma)^4C_\Delta^2,2\left(\eta_X^2+\eta_Y^2\right)^2\right\}=32C_\sigma^4C_\Delta^2$$

for the typical case when $C_\sigma\ge\eta_X,\eta_Y$,

we have, when $0\le x\le c'\tilde{n}^{1/6}$,

$$P\left(\left\{\tilde{n}^{1/6}\left|\langle\widehat{Z^+,Z^+}\rangle_1-\int_0^1\sigma_t^{Z^+2}dt\right|>x\right\}\bigcap\{\Gamma_{C_\sigma}=1\}\right)$$
$$\le 4\exp(-C'x^2),$$

and

$$P\left(\left\{\tilde{n}^{1/6}\left|\langle\widehat{Z^-,Z^-}\rangle_1-\int_0^1\sigma_t^{Z^-2}dt\right|>x\right\}\bigcap\{\Gamma_{C_\sigma}=1\}\right)$$
$$\le 4\exp(-C'x^2),$$

where $c'=4\sqrt{2}C_{1,x}C_2$ and $C'=(32C_2w^2)^{-1}$.

Finally, for the TSCV, when $0\le x\le c\tilde{n}^{1/6}$,

$$P\left(\left\{\tilde{n}^{1/6}\left|\langle\widehat{X,Y}\rangle_1-\int_0^1\sigma_t^{(X)}\sigma_t^{(Y)}\rho_t^{(X,Y)}dt\right|>x\right\}\bigcap\{\Gamma_{C_\sigma}=1\}\right)$$
$$\le P\left(\left\{\tilde{n}^{1/6}\left|\langle\widehat{Z^+,Z^+}\rangle_1-\int_0^1\sigma_t^{Z^+2}dt\right|>2x\right\}\bigcap\{\Gamma_{C_\sigma}=1\}\right)$$
$$+P\left(\left\{\tilde{n}^{1/6}\left|\langle\widehat{Z^-,Z^-}\rangle_1-\int_0^1\sigma_t^{Z^-2}dt\right|>2x\right\}\bigcap\{\Gamma_{C_\sigma}=1\}\right)$$
$$\le 8\exp(-Cx^2),$$

where $c=c'/2=2\sqrt{2}C_{1,x}C_2$ and $C=4C'=(8C_2w^2)^{-1}$. For big enough n and the typical case when $C_\Delta\ge 1$ and $C_\sigma\ge\eta_X$, we have

$$c=4\sqrt{2}C_\sigma^2C_\Delta \quad\text{and}\quad C=\left(256w^2C_\sigma^4C_\Delta^2\right)^{-1}. \qquad(A.29)$$

This completes the proof of the first half of the statement of Theorem 2, when $\Gamma_{C_\sigma}\equiv 1$.

For the second half of the theorem, we have

$$P\left(\tilde{n}^{1/6}\left|\langle\widehat{X,Y}\rangle_1-\int_0^1\sigma_t^{(X)}\sigma_t^{(Y)}\rho_t^{(X,Y)}dt\right|>x\right)$$
$$=P\left(\left\{\tilde{n}^{1/6}\left|\langle\widehat{X,Y}\rangle_1-\int_0^1\sigma_t^{(X)}\sigma_t^{(Y)}\rho_t^{(X,Y)}dt\right|>x\right\}\right.$$
$$\cap\{\Gamma_{C_\sigma}<1\}\bigg)+P(\{\Gamma_{C_\sigma}<1\})$$
$$<8\exp\left(-\left(256w^2C_\sigma^4C_\Delta^2\right)^{-1}x^2\right)+2k_\sigma\exp\left\{-aC_\sigma^b\right\},$$
$$\text{when }0\le x\le 4\sqrt{2}C_\sigma^2C_\Delta\tilde{n}^{1/6}.$$

Let $C_\sigma=(\frac{x^2}{256w^2aC_\Delta^2})^{\frac{1}{4+b}}$, the above inequality becomes

$$P\left\{\tilde{n}^{1/6}\left|\langle\widehat{X,Y}\rangle_1-\int_0^1\sigma_t^{(X)}\sigma_t^{(Y)}\rho_t^{(X,Y)}dt\right|>x\right\}$$
$$\le(8+2k_\sigma)\exp\left\{-(256w^2)^{\frac{-b}{4+b}}\cdot a^{\frac{4}{4+b}}\cdot C_\Delta^{\frac{-2b}{4+b}}\cdot x^{\frac{2b}{4+b}}\right\},\quad(A.30)$$

when $0\le x\le 2^{\frac{5b-12}{2b}}C_\Delta(w^2a)^{\frac{-2}{b}}\cdot\tilde{n}^{\frac{4+b}{6b}}$. $\qquad\square$

Remark 4. Note that the argument is not restricted to TSCV based on the pairwise-refresh times—it works the same (only with $\tilde{n}$ replaced by $\tilde{n}_*$, the observation frequency of the all-refresh method) for the case when the synchronization scheme is chosen to be the all-refresh method, as long as the sampling conditions (Conditions 3–4) are satisfied.

REFERENCES

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010), "High Frequency Covariance Estimates With Noisy and Asynchronous Financial Data," *Journal of the American Statistical Association*, 105, 1504–1517. [412]

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, 351–416. [412]

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008), "Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices With Noise and Non-Synchronous Trading," *Journal of Econometrics*, forthcoming. [412,414,417]

—— (2009), "Realized Kernels in Practice: Trades and Quotes," *Econometrics Journal*, 12, 1–32. [412]

—— (2011), "Subsampling Realised Kernel," *Journal of Econometrics*, 160, 204–219. [412]

Bickel, P. J., and Levina, E. (2008), "Regularized Estimation of Large Covariance Matrices," *The Annals of Statistics*, 36, 199–227. [412]

Fan, J., Fan, Y., and Lv, J. (2008), "Large Dimensional Covariance Matrix Estimation Via a Factor Model," *Journal of Econometrics*, 147, 186–197. [412]

Fan, J., and Wang, Y. (2007), "Multi-Scale Jump and Volatility Analysis for High-Frequency Financial Data," *Journal of the American Statistical Association*, 102, 1349–1362. [412]

Fan, J., Zhang, J., and Yu, K. (2011), "Asset Allocation and Risk Assessment With Gross Exposure Constraints for Vast Portfolios," *Journal of American Statistical Association*, forthcoming. [412,413,414,421]

Hautsch, N., Kyj, R. M., and Oomen, L. C. A. (2009), "A Blocking and Regularization Approach to High Dimensional Realized Covariance Estimation," *Journal of Applied Econometrics*, forthcoming. [412,414]

Hayashi, T., and Yoshida, N. (2005), "On Covariance Estimation of Non-Synchronously Observed Diffusion Processes," *Bernoulli*, 11, 359–379. [412]

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009), "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276. [412]

Jagannathan, R., and Ma, T. (2003), "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 58, 1651–1683. [412,413]

Karatzas, I., and Shreve, S. E. (2000), *Brownian Motion and Stochastic Calculus* (2nd ed.), New York: Springer. [422,426]

Kinnebrock, S., Podolskij, M., and Christensen, K. (2010), "Pre-Averaging Estimators of the Ex-Post Covariance Matrix in Noisy Diffusion Models With Non-Synchronous Data," *Journal of Econometrics*, 159, 116–133. [412]

Lam, C., and Fan, J. (2009), "Sparsistency and Rates of Convergence in Large Covariance Matrices Estimation," *The Annals of Statistics*, 37, 4254–4278. [412]

Ledoit, O., and Wolf, M. (2004), "Honey, I Shrunk the Sample Covariance Matrix," *Journal of Portfolio Management*, 30, 110–119. [413]

Li, Y., and Mykland, P. (2007), "Are Volatility Estimators Robust With Respect to Modeling Assumptions?," *Bernoulli*, 13, 601–622. [412]

Markowitz, H. (1959), *Portfolio Selection: Efficient Diversification of Investments*, New York: Wiley. [412]

Markowitz, H. M. (1952), "Portfolio Selection," *Journal of Finance*, 7, 77–91. [412]

Rothman, A. J., Levina, E., and Zhu, J. (2009), "Generalized Thresholding of Large Covariance Matrices," *Journal of the American Statistical Association*, 104, 177–186. [412]

Tao, M., Wang, Y., Yao, Y., and Zou, J. (2011), "Large Volatility Matrix Inference via Combining Low-Frequency and High-Frequency Approaches," *Journal of the American Statistical Association*, 106, 1025–1040. [412]

Wang, Y., and Zou, J. (2010), "Vast Volatility Matrix Estimation for High-Frequency Financial Data," *The Annals of Statistics*, 38, 943–978. [412]

Xiu, D. (2010), "Quasi-Maximum Likelihood Estimation of Volatility With High Frequency Data," *Journal of Econometrics*, 159, 235–250. [412]

Zhang, L. (2006), "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043. [412]

——— (2011), "Estimating Covariation: Epps Effect and Microstructure Noise," *Journal of Econometrics*, 160, 33–47. [412,414,415]

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005), "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411. [412,415]

Zheng, X., and Li, Y. (2011), "On the Estimation of Integrated Covariance Matrices of High Dimensional Diffusion Processes," *The Annals of Statistics*, 39(6), 3121–3151. [412]