Volatility inference in the presence of both endogenous time and microstructure noise

Yingying Li[a], Zhiyuan Zhang[b,*], Xinghua Zheng[a]

[a] *Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Hong Kong*
[b] *School of Statistics and Management, Shanghai University of Finance and Economics, China*

Abstract

In this article we consider the volatility inference in the presence of both market microstructure noise and endogenous time. Estimators of the integrated volatility in such a setting are proposed, and their asymptotic properties are studied. Our proposed estimator is compared with the existing popular volatility estimators via numerical studies. The results show that our estimator can have substantially better performance when time endogeneity exists.
© 2013 Elsevier B.V. All rights reserved.

MSC: 60F05; 60G44; 62M09

Keywords: Itô process; Realized volatility; Integrated volatility; Time endogeneity; Market microstructure noise

1. Introduction

In recent years there has been growing interest in the inference for asset price volatilities based on high-frequency financial data. Suppose that the latent log price $X = (X_t)$ follows an Itô process

$$dX_t = \mu_t \, dt + \sigma_t \, dW_t, \quad \text{for } t \in [0, 1], \tag{1}$$

where W is a standard Brownian motion, and the drift (μ_t) and volatility (σ_t) are both stochastic processes. Econometric interests are usually in the inference for the integrated volatility, i.e.,

* Corresponding author.
  E-mail addresses: yyli@ust.hk (Y. Li), zhang.zhiyuan@mail.shufe.edu.cn (Z. Zhang), xhzheng@ust.hk (X. Zheng).

quadratic variation, of the log price process

$$\langle X, X \rangle_t = \int_0^t \sigma_s^2 \, ds.$$

A classical estimator from probability theory (see, for example, [16,7]) for this quantity is the realized volatility (RV) based on the discrete time observations

$$X_{t_i} \quad \text{for } 0 = t_0 < t_1 < \cdots < t_{N_1} = 1,$$

where t_i's may be a sequence of stopping times. The RV $[X, X]_t$ is defined as the sum of squared log returns

$$[X, X]_t = \sum_{t_i \leq t} (\Delta X_{t_i})^2,$$

where $\Delta X_{t_i} = X_{t_i} - X_{t_{i-1}}$ for $i \geq 1$. Under mild conditions, when the observation frequency N_1 goes to infinity, $[X, X]_t \xrightarrow{p} \langle X, X \rangle_t$. Furthermore, when the observation times $(t_i)_{i \geq 0}$ are independent of X, a complete asymptotic theory for the estimator $[X, X]_t$ is available, which says that $\sqrt{N_1}([X, X]_t - \langle X, X \rangle_t)$ is asymptotically a mixture of normal whose mixture component is the variance equal to $2 \int_0^t \sigma_s^4 \, dH_s$, where H_t is the "quadratic variation of time" process provided that the following limit exists (see [23] or [25])

$$\text{plim}_{N_1 \to \infty} N_1 \sum_{t_i \leq t} (\Delta t_i)^2 = H_t,$$

where "plim" stands for limit in probability. The quantity $\int_0^t \sigma_s^4 \, dH_s$ can be consistently estimated by the quarticity $N_1/3 \cdot [X, X, X, X]_t := N_1/3 \cdot \sum_{t_i \leq t} (\Delta X_{t_i})^4$.

The above provides a foundation for estimating the integrated volatility based on high frequency data. However, when it comes to the practical side, the assumptions for RV are often violated. Two aspects are of great importance. They are

(a) Market microstructure noise; and
(b) Endogeneity in the price sampling times.

For the first issue, recently there has seen a large literature on estimating quantities of interest with prices observed with microstructure noise. One commonly used assumption is that the noises are additive and one observes

$$Y_{t_i} := X_{t_i} + \varepsilon_{t_i}, \quad \text{for } i = 0, 1, \ldots, N_1. \tag{2}$$

It is often assumed that the noise $(\varepsilon_{t_i})_{i \geq 1}$ is an independent sequence of white noise and the sampling times $(t_i)_{i \geq 1}$ are independent of X. Various estimators of integrated volatility have been proposed. See, for example, two scales realized volatility of Zhang, Mykland and Aït-Sahalia [34], multi-scale realized volatility by Zhang [33], realized Kernels of Barndorff-Nielsen et al. [6], the pre-averaging method by Jacod et al. [15] and the QMLE method by Xiu [31]. Related works include Aït-Sahalia, Mykland and Zhang [4], Bandi and Russell [5], Fan and Wang [9], Hansen and Lunde [13], Kalnina and Linton [17], Li and Mykland [19], Phillips and Yu [26] among others.

In contrast, issue (b) has only recently been brought to researchers' attention. The case when the sampling times are irregular or random but (conditionally) independent of the price process has been studied by Aït-Sahalia and Mykland [3], Duffie and Glynn [8], Meddahi, Renault and

Werker [22], Hayashi, Jacod and Yoshida [14] among others. A recent work of Renault and Werker [28] provides a detailed discussion on the issue of possible endogenous effect that stems from the price sampling times in a semi-parametric context. Li et al. [20] further investigate the time endogeneity effect on volatility estimation in a nonparametric setting. Volatility estimation in the presence of endogenous time in some special situations like when the observation times are hitting times has been studied in [10,12], and in a general situation has also been studied in [11]. In [20], the analysis was carried out by considering the time endogeneity effect which is reflected by

$$\text{plim}\sqrt{N_1}[X, X, X]_t \tag{3}$$

where $\sqrt{N_1}[X, X, X]_t := \sqrt{N_1} \sum_{t_i \le t} (\Delta X_{t_i})^3$ is the tricity. Interestingly, the literature usually neglects the important information one could draw from the quantity $[X, X, X]_t$, which can be interpreted as a measure of the covariance between the price process and time as shown in [20]. Li et al. [20] also conducted empirical work that provides compelling evidence that the endogenous effect does exist in financial data, i.e., $\text{plim}\sqrt{N_1}[X, X, X]_t \ne 0$.

Although individually each issue (a) or (b) has been studied in the literature, there is a lack of studies that take both the microstructure noise and time endogeneity effect into consideration. Robert and Rosenbaum [30] study the estimation of the integrated (co-)volatility for an interesting model where the observation times are triggered by exiting from certain "uncertainty zones", in which case both microstructure noise and time endogeneity may exist. In this paper, we consider the presence of both microstructure noise and time endogeneity in a general setting.

The paper is organized as follows. The setup and assumptions are given in Section 2. The main results are given in Section 3. In Section 4, simulation studies are performed in which our proposed estimator is compared with several existing popular estimators. Section 5 concludes. The proofs (except that of Proposition 1) are given in the Appendix A; the proof of Proposition 1 is given in the supplementary article by Li, Zhang and Zheng [21].

2. Setup and assumptions

Assumption 1. We assume the setting of (1) and (2) and that there is a filtration $(\mathcal{F}_t)_{t \ge 0}$, with respect to which W, μ and σ in (1) are adapted and $(t_i)_{i \ge 1}$ are $(\mathcal{F}_t)$-stopping times. Furthermore, the filtration $(\mathcal{F}_t)$ is generated by finitely many continuous martingales.

In the Introduction, we adopted the notation N_1 for the number of observed prices over time interval $[0, 1]$. Here, we generalize this and denote

$$N_t = \max\{i : t_i \le t\}.$$

In developing limiting results, one should be able to rely on some index variable approaching infinity/zero. In our context, we assume that $\max_i \Delta t_i \xrightarrow{p} 0$ is driven by some underlying force, for instance, $n \to \infty$, where n (non-random) characterizes the sampling frequency over time interval $[0, 1]$.

We aim at effectively estimating $\langle X, X \rangle_t$ based on our general setup. A local averaging approach is adopted. We consider the time endogeneity on the sub-grid level. Take the single sub-grid case for illustration, the sub-sample $\mathcal{S} = \mathcal{S}_0 := \{t_p, t_{p+q}, \dots, t_{p+iq}, \dots\}$ is constructed by choosing every qth observation (starting from the pth observation) from the complete grid.
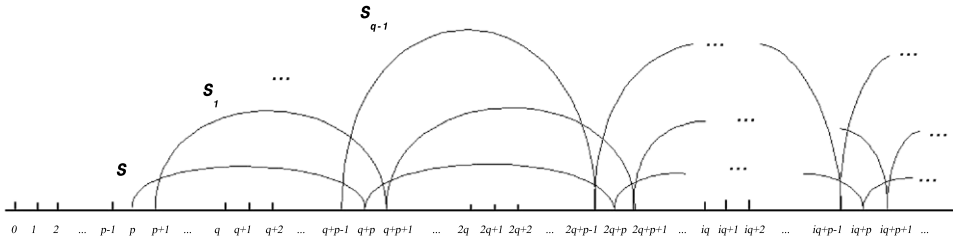
Fig. 1. Grid allocation for local averaging.

Here p is the number of observations that we take in constructing local average, q is the size of blocks, and both are non-random numbers just as n. Define

$$\ell := \left\lfloor \frac{n-p}{q} \right\rfloor, \tag{4}$$

which satisfies that $\ell q \leq n$, and as p shall be taken as $o(n)$, $\ell q/n \to 1$ as $n \to \infty$. As n measures the sampling frequency of the complete grid, ℓ measures that of the sub-grid $\mathcal{S}$. Moreover, for notational ease, for $k = 0, 1, \ldots, q-1$, we define

$$t_{i,j}^k := t_{iq+p-j+k}, \quad \text{for } i = 0, 1, 2, \ldots, \text{ and } j = 0, 1, 2, \ldots, p-1, \tag{5}$$

and let

$$t_{i,j} = t_{i,j}^0.$$

Analogous to (3), we consider the quantity $\sqrt{\ell}[X, X, X]_t^{\mathcal{S}} = \sqrt{\ell} \sum_{t_{i,0} \leq t} (X_{t_{i,0}} - X_{t_{i-1,0}})^3$. The superscript $\mathcal{S}$ indicates the calculation being performed is based on the designated sub-grid. This convention applies to other sub-grids. Moving the sub-grid $\mathcal{S}$ one step forward forms sub-grid $\mathcal{S}_1$, continuing this process gives sub-grid $\mathcal{S}_2$ and so on till the $(q-1)$th sub-grid $\mathcal{S}_{q-1}$. Fig. 1 provides a graphical demonstration of our grid allocation. Further, on the sub-grid $\mathcal{S}$, we define the number of observations up to time t in sub-grid $\mathcal{S}$ as

$$L_t := \max\{i : t_{i,0} \leq t\}.$$

Naturally, L_1 and N_1 satisfy $L_1 \leq N_1/q$.

3. Main results

We start with results based on a single sub-grid and then proceed to the multiple sub-grids case.

3.1. Single sub-grid: local averaging

A natural and effective way of reducing the effect of microstructure noise in estimating $\langle X, X \rangle_t$ is averaging; see, e.g., [15,27]. Following Jacod et al. [15], we average every p observations that precede each observation in the sub-sample $\mathcal{S}$ to obtain a new sequence of observations, which we denote by $(\overline{Y}_{t_{i,0}})_{i \geq 0}$. Based on this sequence of observations, we obtain a single-grid biased local averaging estimator. To be specific,

$$\overline{Y}_{t_{i,0}} = \frac{1}{p} \sum_{j=0}^{p-1} Y_{t_{i,j}}, \quad \text{for } i = 0, 1, 2, \ldots.$$

The RV based on the $\overline{Y}$ sequence is denoted by

$$[\overline{Y}, \overline{Y}]_t^{\mathcal{S}} = \sum_{t_{i,0} \leq t} \left(\Delta \overline{Y}_{t_{i,0}}\right)^2,$$

where $\Delta \overline{Y}_{t_{i,0}} = \overline{Y}_{t_{i,0}} - \overline{Y}_{t_{i-1,0}}$ for $i \geq 1$. After correcting the bias due to noise, the single-grid local averaging estimator is defined as

$$\widehat{\langle X, X \rangle}_t^{LA} := [\overline{Y}, \overline{Y}]_t^{\mathcal{S}} - \frac{2L_t}{p} \widehat{\sigma_\varepsilon^2}, \quad \text{for } t \in [0, 1], \tag{6}$$

where

$$\widehat{\sigma_\varepsilon^2} = [Y, Y]_1 / 2N_1, \tag{7}$$

is an estimator of σ_ε^2, see Lemma 1 in Appendix A.1, and $[Y, Y]_1$ is the RV based on all observations up to time 1. We now state conditions that lead to the theorem for the single sub-grid case:

C(1). μ_t and $\sigma_t^2 \geq c > 0$ are integrable and locally bounded;
C(2). $n/N_1 = O_p(1)$;
C(3). $\Delta_n := \max_{1 \leq i \leq N_1} |t_i - t_{i-1}| = O_p(1/n^{1-\eta})$ for some nonnegative constant η;
C(4). $L_t/\ell \xrightarrow{p} \int_0^t r_s ds$ in $D[0, 1]$, where r_s is an adapted integrable process (and hence in particular, $N_1/n = O_p(1)$);
C(5). the microstructure noise sequence $(\varepsilon_{t_i})_{i \geq 0}$ consists of independent random variables with mean 0, variance σ_ε^2, and common finite third and forth moments, and is independent of $\mathcal{F}_1$.

The following theorem characterizes the asymptotic property of the estimator (6).

Theorem 1. *Assume Assumption* 1 *and conditions* C(1)–C(5) . *Suppose that* $\eta \in [0, 1/6)$*, and* $\ell \sim C_\ell n^\alpha$ *and* $p \sim C_p n^\alpha$ *for some* $0 < \alpha < 2(1 - \eta)/5$ *and positive constants* C_ℓ *and* C_p*, and also that*

$$\ell[X, X, X, X]_t^{\mathcal{S}} \xrightarrow{p} \int_0^t u_s \sigma_s^4 \, ds \quad \text{for every } t \in [0, 1], \text{ and} \tag{8}$$

$$\sqrt{\ell}[X, X, X]_t^{\mathcal{S}} \xrightarrow{p} \int_0^t v_s \sigma_s^3 \, ds \quad \text{for every } t \in [0, 1], \tag{9}$$

where $[X, X, X, X]_t^{\mathcal{S}} = \sum_{t_{i,0} \leq t} (X_{t_{i,0}} - X_{t_{i-1,0}})^4$*, and* $u_s \sigma_s^4$ *and* $v_s^2 \sigma_s^4$ *are both integrable. Then, stably in law,*

$$\sqrt{\ell} \left(\widehat{\langle X, X \rangle}_t^{LA} - \langle X, X \rangle_t \right) \Longrightarrow \underbrace{\frac{2}{3} \int_0^t v_s \sigma_s dX_s + \int_0^t \left[\left(\frac{2}{3} u_s - \frac{4 v_s^2}{9} \right) \sigma_s^4 \right.}_{\text{asymptotic bias}}$$

$$\left. + 12 r_s \left(\frac{C_\ell}{C_p} \sigma_\varepsilon^2 \right)^2 + 8 \frac{C_\ell}{C_p} \sigma_s^2 \sigma_\varepsilon^2 \right]^{1/2} dB_s$$

where B_t *is a standard Brownian motion independent of* $\mathcal{F}_1$.

Proof of the theorem is given in Appendix A.2.

In the literature it is often assumed that the mesh $\Delta_n = O_p(1/n)$, in other words, $\eta = 0$ in Condition C(4). In this case, the convergence rate in Theorem 1 can be arbitrarily close to $n^{1/5}$.

Remark 1. Unlike in the full grid setting where a nonzero limit of tricity can be easily generated by letting the sampling times be hitting times of asymmetric barriers (see for instance Examples 4 and 5 of Li et al. [20]), in the subgrid case a nonzero limit of tricity is far less common, and in particular under the settings of both Examples 4 and 5 of Li et al. [20], the limit in (9) vanishes. However as we found in simulation studies (not all reported), even in these situations, adopting the (finite sample) bias correction discussed in Section 3.3 can substantially reduce the (finite sample) bias. Similar remark applies to the estimator in Theorem 2.

3.2. Multiple sub-grids: moving average

We show in this subsection that for any $\varepsilon > 0$, rate $n^{1/4-\varepsilon}$ consistency can be achieved by using moving average based on multiple sub-grids. For that purpose, we need such notations as $[\overline{Y}, \overline{Y}]_t^{\mathcal{S}_k}$, i.e. the RV of locally averaged Y process over the kth sub-grid, for the same operations that are performed over the 0th sub-grid $\mathcal{S} = \mathcal{S}_0$ being adjusted to the kth sub-grid $\mathcal{S}_k$. To be specific, we take $[\overline{Y}, \overline{Y}]_t^{\mathcal{S}_k}$ for example; other notation with superscript k or $\mathcal{S}_k$ has similar interpretation. Similar to the definition of $[\overline{Y}, \overline{Y}]_t^{\mathcal{S}}$ (i.e. $[\overline{Y}, \overline{Y}]_t^{\mathcal{S}_0}$), we first define

$$\overline{Y}_{t_{i,0}^k} := \frac{1}{p} \sum_{j=0}^{p-1} Y_{t_{iq+p+k-j}}, \quad \text{for } i = 0, 1, 2, \ldots,$$

where, recall that $t_{i,0}^k = t_{iq+p+k}$ denotes the ith observation time on the kth sub-grid. The RV of locally averaged Y process over the kth sub-grid is defined as follows

$$[\overline{Y}, \overline{Y}]_t^{\mathcal{S}_k} := \sum_{t_{i,0}^k \leq t} (\Delta \overline{Y}_{t_{i,0}^k})^2,$$

where $\Delta \overline{Y}_{t_{i,0}^k} = \overline{Y}_{t_{i,0}^k} - \overline{Y}_{t_{i-1,0}^k}$ for $i \geq 1$. Assume the following conditions that lead to the asymptotic result on multiple sub-grids:

C(6). $\ell \sum_{t_i \leq t} \left(\sum_{j=1}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j}} \right)^2 (\Delta X_{t_i})^2 \xrightarrow{p} \int_0^t w_s \sigma_s^4 \, ds$ for every $t \in [0, 1]$, where $w_s \sigma_s^4$ is integrable;

C(7). $\frac{1}{q} \sum_{k=0}^{q-1} \sqrt{\ell}[X, X, X]_t^{\mathcal{S}_k} \xrightarrow{p} \int_0^t \bar{v}_s \sigma_s^3 \, ds$ for every $t \in [0, 1]$, where $\bar{v}_s^2 \sigma_s^4$ is integrable.

Define

$$A(p, q) := \frac{2}{q} \sum_{j=1}^{p-1} \left(\frac{j^2}{p^2} - \frac{j}{p} \right).$$

Under the conditions of Theorem 2, $A(p, q) \sim -n^{4\alpha-2} C_\ell C_p / 3$.

Theorem 2. *Assume Assumption 1 and conditions* C(1)–C(7). *Suppose that $\eta \in [0, 1/9)$, and $\ell \sim C_\ell n^\alpha$ and $p \sim C_p n^{3\alpha-1}$ for some $\max(4\eta, 1/3) < \alpha < (1 - \eta)/2$ and positive constants*

C_ℓ and C_p. Then, stably in law,

$$\sqrt{\ell}\left(\frac{1}{q}\sum_{k=0}^{q-1}[\overline{Y},\overline{Y}]_t^{\mathcal{S}_k} - \frac{2N_t}{pq}\widehat{\sigma_\varepsilon^2} - (1+A(p,q))\langle X,X\rangle_t\right)$$

$$\Longrightarrow \underbrace{\frac{2}{3}\int_0^t \bar{v}_s\sigma_s\,dX_s}_{\text{asymptotic bias}} + \int_0^t\left[\left(4w_s - \frac{4}{9}\bar{v}_s^2\right)\sigma_s^4 + \frac{8C_\ell^3}{C_p}r_s(\sigma_\varepsilon^2)^2\right]^{1/2}dB_s,$$

where B_t is a standard Brownian motion that is independent of $\mathcal{F}_1$.

Proof of the theorem is given in Appendix A.3.

If one assumes that $\Delta_n = O_p(1/n)$, then $\eta = 0$, and the convergence rate in the above theorem can be arbitrarily close to $n^{1/4}$.

Remark 2. If times are exogenous, Condition C(6) can be reduced to a similar assumption as (48) on p. 1401 of Zhang, Mykland and Aït-Sahalia [34]. The limit is then related to quarticity and can be consistently estimated; see, e.g., [15,6]. In general, when observation times can be endogenous, the limit is expected to be different.

3.3. Bias correction

Since the estimator constructed based on multiple grids achieves a better rate of convergence, below we shall mainly focus on the moving average setting. Based on the above result, we have the following (infeasible) unbiased estimator:

$$\widehat{\langle X,X\rangle}_1^{(0)} := \frac{-\frac{2}{3}\int_0^1 \bar{v}_s\sigma_s\,dX_s + \sqrt{\ell}\left(\frac{1}{q}\sum_{k=0}^{q-1}[\overline{Y},\overline{Y}]_1^{\mathcal{S}_k} - \frac{2N_1}{pq}\widehat{\sigma_\varepsilon^2}\right)}{\sqrt{\ell}(1+A(p,q))}.$$

The following corollary describes the asymptotic property for this estimator.

Corollary 1. *Under the assumptions of Theorem 2, stably in law,*

$$\sqrt{\ell}\left(\widehat{\langle X,X\rangle}_1^{(0)} - \langle X,X\rangle_1\right) \Longrightarrow \int_0^1\left[\left(4w_s - \frac{4}{9}\bar{v}_s^2\right)\sigma_s^4 + \frac{8C_\ell^3}{C_p}r_s(\sigma_\varepsilon^2)^2\right]^{1/2}dB_s,$$

where B_t is a standard Brownian motion that is independent of $\mathcal{F}_1$.

Proof. This is just a rearrangement of the convergence in Theorem 2. $\square$

To improve over $\widehat{\langle X,X\rangle}_1^{(0)}$ and build a feasible unbiased estimator, a consistent estimator for the bias term $2/3\int_0^t \bar{v}_s\sigma_s\,dX_s$ is needed. This is the issue that we deal with next. Define

$$F_n^{(2)}(t) := \frac{\sqrt{\ell}\left(\frac{1}{q}\sum_{k=0}^{q-1}[\overline{Y},\overline{Y}]_t^{\mathcal{S}_k} - \frac{2N_t}{pq}\widehat{\sigma_\varepsilon^2}\right)}{\sqrt{\ell}(1+A(p,q))}, \quad \text{and} \quad f^{(2)}(t) := \sigma_t^2, \tag{10}$$

$$F_n^{(3)}(t) := \frac{1}{q}\sum_{k=0}^{q-1}\sqrt{\ell}[\overline{Y},\overline{Y},\overline{Y}]_t^{\mathcal{S}_k}, \quad \text{and} \quad f^{(3)}(t) := \bar{v}_t\sigma_t^3.$$

For a given partition $(\tau_i)_{i \geq 0}$ over $[0, 1]$, we define

$$f_n^{(j)}(t) := (F_n^{(j)}(\tau_i) - F_n^{(j)}(\tau_{i-1}))/(\tau_i - \tau_{i-1}), \quad \text{for } t \in [\tau_i, \tau_{i+1}), \text{ for } j = 2, 3. \quad (11)$$

We then have that stably in law,

$$\sqrt{\ell} \left(F_n^{(2)}(t) - \langle X, X \rangle_t \right) \Longrightarrow \frac{2}{3} \int_0^t \bar{v}_s \sigma_s dX_s + \int_0^t \left[\left(4w_s - \frac{4}{9} \bar{v}_s^2 \right) \sigma_s^4 \right.$$
$$\left. + \frac{8C_\ell^3}{C_p} r_s (\sigma_\varepsilon^2)^2 \right]^{1/2} dB_s.$$

Define

$$\gamma(\alpha, \eta) := \min\{-2\alpha + 1 - 3\eta/2; \alpha/2 - \eta/2; 7\alpha/2 - 3/2; 3\alpha/2 - 1/2 - \eta;$$
$$5\alpha/2 - 1 - \eta/2\}.$$

And assume

C(7') $\left| \frac{1}{q} \sum_{k=0}^{q-1} \sqrt{\ell} [X, X, X]_t^{\mathcal{S}_k} - \int_0^t \bar{v}_s \sigma_s^3 ds \right| / \delta_n \xrightarrow{p} 0$ in $D[0, 1]$ for a (nonrandom) sequence $(\delta_n)_{n \geq 1}$ with $\delta_n \to 0$ and $1/\delta_n = o\left(n^{\gamma(\alpha, \eta)}\right)$.

We have the following.

Proposition 1. *Assume the conditions of Theorem 2,* C(7') *and* $3/7 < \alpha < (2 - 3\eta)/4$ *with* $\eta \in [0, 2/21)$. *Suppose* $f^{(j)}(t)$ *is a.s. continuous and bounded on* $[0, 1]$ *for* $j = 2, 3$. *Moreover, define a partition* $[\tau_i, \tau_{i+1}] := [t_{id_1q}, t_{(i+1)d_1q}]$ *which is a block of* d_1q *time intervals over the complete grid with* $1/d_1 = o\left(1/n^{1-2\alpha}\right)$, $\max_i |\tau_i - \tau_{i-1}| = o_p(1)$ *and* $\delta_n / \min_i |\tau_i - \tau_{i-1}| = O_p(1)$; *and let*

$$\Delta \overline{Y}_{\tau_i} := \frac{1}{p} \sum_{j=0}^{p-1} Y_{t_{id_1q+p-j}} - \frac{1}{p} \sum_{j=0}^{p-1} Y_{t_{(i-1)d_1q+p-j}}.$$

Then

$$\sum_{\tau_i \leq t} \frac{f_n^{(3)}(\tau_{i-1})}{f_n^{(2)}(\tau_{i-1})} \Delta \overline{Y}_{\tau_i} \xrightarrow{p} \int_0^t \bar{v}_s \sigma_s \, dX_s \quad \text{in } D[0, 1].$$

Proof of Proposition 1 is given in the supplementary article of Li, Zhang and Zheng [21].

According to the above proposition, a consistent estimator for the bias $2/3 \int_0^1 \bar{v}_s \sigma_s \, dX_s$ is given by

$$\mathcal{B} := \frac{2}{3} \sum_{\tau_i \leq 1} \frac{f_n^{(3)}(\tau_{i-1})}{f_n^{(2)}(\tau_{i-1})} \Delta \overline{Y}_{\tau_i}.$$

Finally, we define our feasible unbiased estimator as

$$\widehat{\langle X, X \rangle}_1 := \frac{-\mathcal{B} + \sqrt{\ell} \left(\frac{1}{q} \sum_{k=0}^{q-1} [\overline{Y}, \overline{Y}]_1^{\mathcal{S}_k} - \frac{2N_1}{pq} \widehat{\sigma_\varepsilon^2} \right)}{\sqrt{\ell}(1 + A(p, q))}.$$

The following theorem gives the CLT for our final estimator.

Theorem 3. *Under the assumptions of Theorem 2 and Proposition 1, stably in law,*

$$\sqrt{\ell} \left(\widehat{\langle X, X \rangle}_1 - \langle X, X \rangle_1 \right) \Longrightarrow \int_0^1 \left[\left(4w_s - \frac{4}{9} \bar{v}_s^2 \right) \sigma_s^4 + \frac{8C_\ell^3}{C_p} r_s (\sigma_\varepsilon^2)^2 \right]^{1/2} dB_s,$$

where B is a standard Brownian motion independent of $\mathcal{F}_1$.

4. Simulation studies

In this section, we conduct simulation studies. We investigate the performance of our proposed estimator $\widehat{\langle X, X \rangle}_1$ compared with existing popular estimators in both endogenous and non-endogenous cases. We shall use two data generating mechanisms for X: (1) a constant volatility Brownian bridge and (2) a stochastic volatility Heston bridge. In each case, we start the latent process X at $X_0 = \log(5)$, let the standard deviation of the noise be $\sigma_\varepsilon := (\sigma_\varepsilon^2)^{1/2} = 0.0005$ and simulate 1000 sample paths for observed price process Y.

4.1. Estimators used for comparison

Below we briefly recall four commonly used volatility estimators: the two scales realized volatility (TSRV) of Zhang, Mykland and Aït-Sahalia [34], the multi-scale realized volatility (MSRV) of Zhang [33], the realized Kernel estimator of Barndorff-Nielsen et al. [6], and the pre-averaging estimator of Jacod et al. [15].

The (small-sample adjusted) TSRV estimator is given by

$$\widehat{\langle X, X \rangle}_1^{\text{tsrv}} = \left(1 - \frac{1}{K_{\text{tsrv}}} \right)^{-1} \left(\frac{1}{K_{\text{tsrv}}} \sum_{k=1}^{K_{\text{tsrv}}} [Y, Y]_1^{(k)} - \frac{1}{K_{\text{tsrv}}} [Y, Y]_1 \right),$$

where the data is divided into K_{tsrv} non-overlapping sub-grids and $[Y, Y]_1^{(k)}$ is the RV on the kth sub-grid. Zhang, Mykland and Aït-Sahalia [34] provided a guideline on the choice of the grid allocation. If we pretend that the volatility were constant, then the optimal choice for grid allocation is $K_{\text{tsrv}} = c_{\text{tsrv}} N_1^{2/3}$, where, in practice, one can set $c_{\text{tsrv}} = \left(\frac{12([Y,Y]_1/(2N_1))^2}{([Y,Y]_1^{\text{sub}})^2} \right)^{1/3}$, where $[Y, Y]_1^{\text{sub}}$ is the RV based on sparse sampling. Here, we implement $[Y, Y]_1^{\text{sub}}$ at 5 min frequency.

The MSRV estimator, which is a rate-optimal extension to TSRV, is given as follows

$$\widehat{\langle X, X \rangle}_1^{\text{msrv}} = \sum_{j=1}^{K_{\text{msrv}}} \lambda_j \frac{1}{j} \sum_{k=1}^{j} [Y, Y]_1^{(k)},$$

where $\lambda_1 = a_1 + ((N_1 + 1)/2)^{-1}$, $\lambda_2 = a_2 - ((N_1 + 1)/2)^{-1}$ and $\lambda_i = a_i$ for $i \geq 3$ with $a_i = h(i/K_{\text{msrv}})i/K_{\text{msrv}}^2 - h'(i/K_{\text{msrv}})i/(2K_{\text{msrv}}^3)$, for $i = 1, \ldots, K_{\text{msrv}}$, where $K_{\text{msrv}} = c_{\text{msrv}} N_1^{1/2}$ and $h(x) = 12x - 6$. The optimal choice of c_{msrv} when the volatility is constant is

$$c_{\text{msrv}} = \left(\frac{T_3 + T_4 + \left((T_3 + T_4)^2 + 12T_1 T_2 \right)^{1/2}}{2T_2} \right)^{1/2},$$

where $T_1 = 48([Y, Y]_1/(2N_1))^2$, $T_2 = 52([Y, Y]_1/(2N_1))^2/35$, $T_3 = 24([Y, Y]_1/(2N_1))^2/5$ and $T_4 = 48[Y, Y]_1^{\text{sub}}([Y, Y]_1/(2N_1))/5$.

The realized Kernel estimator is defined as

$$\widehat{\langle X, X \rangle}_1^{\text{Ker}} = [Y, Y]_1 + \sum_{h=1}^{H} f_k((h-1)/H)\left[\sum_{i=1}^{N_1} \left(\Delta Y_{t_i} \Delta Y_{t_{i-h}} + \Delta Y_{t_i} \Delta Y_{t_{i+h}}\right)\right],$$

where $H = c_{\text{ker}} N_1^{1/2}$ and f_k is a Kernel function. We choose the Parzen Kernel:

$$f_k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & \text{for } 0 \le x \le 1/2; \\ 2(1-x)^3 & \text{for } 1/2 \le x \le 1. \end{cases}$$

Under constant volatility, the optimal choice for c_{ker} in practice is given by

$$c_{\text{ker}} = \left(\frac{[Y, Y]_1/(2N_1)}{[Y, Y]_1^{\text{sub}}}\right)^{1/2} \left(\frac{1}{f_k^{0,0}}\left(-f_k^{0,2} + \left((f_k^{0,2})^2\right.\right.\right.$$
$$\left.\left.\left. + 3f_k^{0,0}(f_k'''(0) + f_k^{0,4})\right)^{1/2}\right)\right)^{1/2},$$

where $f_k^{0,0} = \int_0^1 f_k(x)^2 dx$, $f_k^{0,2} = \int_0^1 f_k(x)f_k''(x)dx$ and $f_k^{0,4} = \int_0^1 f_k(x)f_k'''(x)dx$.

The pre-averaging estimator is as follows:

$$\widehat{\langle X, X \rangle}_1^{\text{Pre}} = \frac{1}{\theta\varphi_2\sqrt{N_1}} \sum_{i=0}^{N_1-k_n+1} (\Delta\widehat{Y}_i)^2 - \frac{\varphi_1}{2\theta^2\varphi_2 N_1}[Y, Y]_1,$$

where $\varphi_1 = 1$, $\varphi_2 = 1/12$ and

$$\Delta\widehat{Y}_i = \frac{1}{k_n}\left(\sum_{j=k_n/2}^{k_n-1} Y_{i+j} - \sum_{j=0}^{k_n/2-1} Y_{i+j}\right)$$

with $k_n = \sqrt{N_1}\theta$. The optimal choice of θ when the volatility is constant is

$$\theta = 4.777([Y, Y]_1/(2N_1))^{1/2}/([Y, Y]_1^{\text{sub}})^{1/2}.$$

Remark 3. The grid allocation schemes in constructing the above estimators are optimal in the sense of achieving efficient asymptotic variance bound when (σ_t) is constant. However, in practice there is no optimal choice since, for instance, (σ_t) is random and time dependent. See Remarks 2 and 3 in [15] for related discussions on this. In our case, due to the more complex model assumptions, i.e. data with time endogeneity and noise, and grid allocation scheme, i.e. bivariate setting (p, q) in contrast to the existing univariate cases, we do not provide a theoretical optimal choice but rather give below some practical guidelines.

Back to our estimator $\widehat{\langle X, X \rangle}_1$, there are several tuning parameters (n, ℓ, p, q and d_1) that one has to determine. Regarding n which characterizes the sampling frequency, one can use the average number of transactions per day for the past, say 30, days as an approximation. About (ℓ, p, q), notice that Theorem 2 suggests $\ell \sim C_\ell n^\alpha$ (hence $q \sim n/\ell$) and $p \sim C_p n^{3\alpha-1}$. On the one hand, one should choose ℓ as large as possible in order to have higher convergence rate. On the other hand, large ℓ induces small q and hence small p (recall $q > p$) and the main role that

p plays is to reduce the microstructure noise. Hence, one should also be aware of the magnitude of the microstructure noise when choosing appropriate p, and p cannot be too small when prices are heavily contaminated. Under the simulation setting below, the sampling frequency is around $n = 46,800$, and the standard deviation of the noise is $\sigma_\varepsilon = 0.0005$. We choose $p = 5$ which is found to be good enough to reduce the microstructure noise effect. In practice, one can use (7) to estimate the standard deviation of the noise and come up with a reasonable choice of p. The block size q should be larger than p and is chosen as 20 (and $\ell \approx 2340$). As to d_1, this depends on, for example, how volatile the volatility process is, which one can get some rough idea by looking at a suitable estimate of the spot volatilities. If the volatility process is more volatile, one should divide the whole time interval into shorter time periods, i.e., choose a smaller d_1. In our simulation, we choose $d_1 = 100$, i.e. dividing the complete grid into around 20 blocks.

We next present our three simulation designs and the corresponding results.

4.2. Design I: Brownian bridge with hitting times

We first consider the case when the latent price process X follows a Brownian bridge with (constant) volatility σ that starts at X_0 and ends at $X_0 + 4\sigma$. X can be expressed as (see p. 358 of Karatzas and Shreve [18])

$$dX_t = \frac{X_0 + 4\sigma - X_t}{1 - t} \, dt + \sigma \, dW_t,$$

where W_t is a standard Brownian motion. In this study, we set $\sigma = 0.02$. The sampling times are generated as follows: let $a = 5\sigma, b = \sigma/10, n = 46,800, \ell' \approx 16\,800$ (roughly $n^{19/21}$), and $q' = [n/\ell']$. Then

(1) for $j = 0, 1, 2, \ldots, q', \ t_j = \frac{j}{2n}$;
(2) for $i = 1, 2, \ldots,$
    sparse sampling: $t_{iq'+1} = \inf\{t > t_{iq'} : X_t - X_{t_{iq'}} = \text{ either } a/\sqrt{\ell'} \text{ or } -b/\sqrt{\ell'}\}$;
    intensive sampling: $t_{iq'+j} = t_{iq'+1} + \frac{j-1}{2n}$, for $j = 2, \ldots, q'$.

The mean observation duration when sampling sparsely is about $1/(2\ell')$, roughly 3 times of the observation duration when sampling intensively. If as $n \to \infty, \ell'$ grows in the rate of $n^{19/21}$, then actually the limit in C(7) vanishes, however, as one can see from the simulation results below, (finite sample) bias correction as discussed in Section 3.3 can substantially reduce the (finite sample) bias.

Fig. 2 displays the histogram and normal Q–Q plot for the estimator $\widehat{\langle X, X \rangle}_1$ based on the 1000 simulated samples. The plots show that the finite sample behavior of our CLT works well. In Table 1 we compare the performances of the four estimators that we discussed in Section 4.1, the "Uncorrected" estimator $F_n^{(2)}(1)$ defined in (10), and our final estimator $\widehat{\langle X, X \rangle}_1$. From the table one can see that our estimator provides the smallest RMSE and has substantially smaller bias than the others (reduced by more than 80%) while maintains similar efficiency (standard deviation).

4.3. Design II: Heston bridge with hitting times

In order to further investigate the performance of our estimator under more complex situations, in this subsection, we consider the following stochastic volatility model

$$\begin{cases} dX_t = \dfrac{X_0 + 4\vartheta^{1/2} - X_t}{1 - t} \, dt + \sqrt{V_t} \, dW_t \\ dV_t = \kappa(\vartheta - V_t) \, dt + \gamma \sqrt{V_t} \, dW_t^\sigma, \end{cases}$$
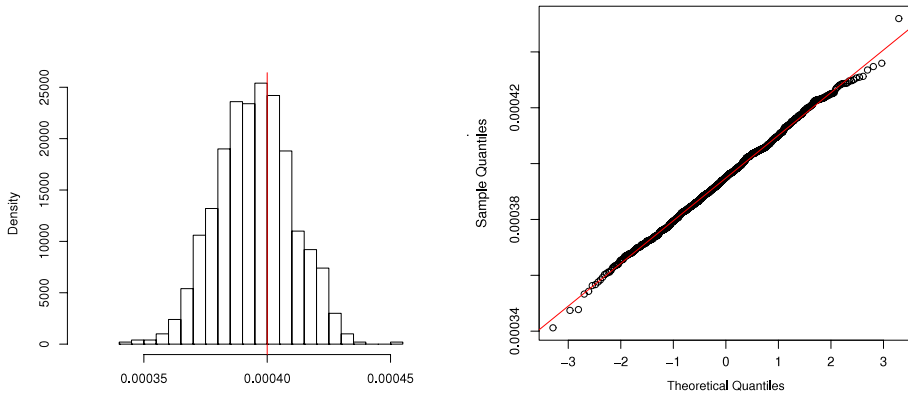
Fig. 2. Histogram and QQ plot of the estimator $\widehat{\langle X, X \rangle}_1$ for Design I. The red vertical line in the histogram indicates the true value of target. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Performance of the six estimators in the presence of endogenous time for Design I, the constant volatility case. Our estimator $\widehat{\langle X, X \rangle}_1$ provides the smallest RMSE. The RMSE is reduced by more than 50%; the bias is reduced by more than 80% while the standard deviation is kept at the same level as others.

	TSRV	MSRV	Kernel	Pre-averaging	Uncorrected	$\widehat{\langle X, X \rangle}_1$
RMSE	3.734e−05	3.553e−05	3.810e−05	3.340e−05	3.300e−05	1.621e−05
Sample bias	3.300e−05	3.163e−05	3.454e−05	2.927e−05	2.911e−05	−4.997e−06
Sample s.d.	1.748e−05	1.619e−05	1.609e−05	1.609e−05	1.555e−05	1.543e−05

Table 2
Performance of the six estimators in the presence of endogenous time for Design II, the stochastic volatility case. Our estimator again provides the smallest RMSE. The RMSE is reduced by more than 50%; the bias is reduced by more than 80%.

	TSRV	MSRV	Kernel	Pre-averaging	Uncorrected	$\widehat{\langle X, X \rangle}_1$
RMSE	3.824e−05	3.579e−05	3.835e−05	3.387e−05	3.375e−05	1.636e−05
Sample bias	3.393e−05	3.175e−05	3.463e−05	2.965e−05	2.974e−05	−4.215e−06

where W_t and W_t^σ are standard Brownian motions with instantaneous correlation coefficient ρ, and κ, ϑ and γ are positive constants. We consider the situation when X starts at X_0 and ends at $X_0 + 4\vartheta^{1/2}$. In the simulation, we set $\vartheta = 0.0004$, $\gamma = 0.5/252$, $\kappa = 5/252$ and $\rho = -0.5$. Here, we choose a moderate value -0.5 for ρ to represent the leverage effect. The leverage effect can be bigger for indices as studied by Aït-Sahalia and Kimmel [2] and Aït-Sahalia et al. [1]. Times are generated according to the same hitting rule as in Design I. We can see from Table 2 that in this more complex situation, our estimator again has substantially smaller bias and RMSE than the others. We did not include the sample standard deviation here since the integrated volatility to be estimated in this case depends on the sample path and is random.

4.4. Design III: Brownian bridge with independent Poisson times

The goal of this design is to check the performance of our estimator when the sampling times are not endogenous. We again assume the Brownian bridge dynamic for X as in Design I. The

Table 3
Performance of the six estimators when the observation times are *not* endogenous. The performance of our estimator is comparable to others.

	TSRV	MSRV	Kernel	Pre-averaging	Uncorrected	$\widehat{\langle X, X \rangle}_1$
RMSE	1.486e−05	1.375e−05	1.434e−05	1.373e−05	1.312e−05	1.568e−05
Sample bias	2.643e−06	1.584e−06	4.144e−06	−2.847e−07	−1.274e−06	−7.723e−06
Sample s.d.	1.463e−05	1.367e−05	1.374e−05	1.373e−05	1.307e−05	1.365e−05

observation times are now generated from an independent Poisson process with rate 46,800. Table 3 reports the result of performance comparison, and we can see that our estimator performs similarly as the other estimators in this case.

In summary, one observes from Tables 1–3 that when sampling times are endogenous (Designs I and II), one can have substantial reductions in RMSE and bias by using our estimator. When there is no endogeneity (Design III), our estimator performs comparably to others.

5. Concluding remarks

In this paper, we establish a theoretical framework for dealing with effects of both the endogenous time and microstructure noise in volatility inference. An estimator that can accommodate both issues is proposed. Numerical studies are performed. The results show that our proposed estimator can substantially outperform existing popular estimators when time endogeneity exists, while has a comparable performance to others when there is no endogeneity.

Acknowledgments

Appendix A. Proofs

Throughout the proofs, C, c, C_1, etc. denote generic constants whose values may change from line to line. Moreover, since we shall establish stable convergence, by a change of measure argument (see e.g. Proposition 1 of Mykland and Zhang [25]) we can suppress the drift and assume that

1. $\mu_t \equiv 0$.
   Moreover, because of the local boundedness condition on σ_t^2, by standard localization arguments we can assume without loss of generality that
2. $0 < c \leq \sigma_t \leq \sigma_+$, where c and σ_+ are nonrandom numbers,
   see e.g. [24,25]. Similarly, we can without loss of generality strengthen the assumption on Δ_n and N_1 in C(2)–C(4) as follows:
3. $\Delta_n \leq C/n^{1-\eta}$, and
4. $n/C \leq N_1 \leq Cn$.

A.1. Prerequisites

In the proofs, we shall repeatedly use the following inequalities.

Burkholder–Davis–Gundy (BDG) inequality with random times:

First, if t_i's are stopping times and $f(s)$ is adapted with $\max_{0 \le s \le 1} |f(s)| \le f_+$, then by the Burkholder–Davis–Gundy inequality with random times (see, e.g., p. 161 of Revuz and Yor [29]), for any exponent $\beta \ge 1$,

$$E \left(\int_{t_{i-1}}^{t_i} f(s) dW_s \right)^\beta \le CE \left(\int_{t_{i-1}}^{t_i} f(s)^2 ds \right)^{\beta/2} \le C f_+^\beta E(t_i - t_{i-1})^{\beta/2}.$$

Doob's L^p inequality:

Second, for any process Z, which is either a continuous time martingale or a positive submartingale, Doob's L^p inequality (see p. 54 of Revuz and Yor [29]) states that, for any $\beta \ge 1$ and any $\lambda > 0$,

$$P \left[\sup_{s \in [0,1]} |Z_s| \ge \lambda \right] \frac{1}{\lambda^\beta} E|Z_1|^\beta,$$

and for $\beta > 1$,

$$\left(E \left[\sup_{s \in [0,1]} |Z_s| \right]^\beta \right)^{1/\beta} \le \frac{\beta}{\beta - 1} \left(E|Z_1|^\beta \right)^{1/\beta}.$$

Therefore, if we can establish a bound order for $E|Z_1|^\beta$ ($\beta = 1$ or 2 in our case), then the same bound order applies in $D[0, 1]$.

We will also use the following results about the convergence of $\widehat{\sigma_\varepsilon^2}$ to σ_ε^2.

Lemma 1. *For $\widehat{\sigma_\varepsilon^2}$ defined in* (7), *one has* $\sqrt{N_1} \left(\widehat{\sigma_\varepsilon^2} - \sigma_\varepsilon^2 \right) = O_p(1)$.

Proof. First, notice that

$$\sqrt{N_1} \left(\widehat{\sigma_\varepsilon^2} - \sigma_\varepsilon^2 \right) = [X, X]_1 / 2\sqrt{N_1} + [X, \varepsilon]_1 / \sqrt{N_1} + ([\varepsilon, \varepsilon]_1 - 2N_1 \sigma_\varepsilon^2) / 2\sqrt{N_1}.$$

By C(2) and the fact that $[X, X]_1 = O_p(1)$,

$$[X, X]_1 / 2\sqrt{N_1} = O_p \left(\frac{1}{\sqrt{n}} \right).$$

As to $[X, \varepsilon]_1 / \sqrt{N_1}$, we treat it as follows,

$$[X, \varepsilon]_1 / \sqrt{N_1} = \frac{1}{\sqrt{N_1}} \sum_{t_i \le 1} \Delta X_{t_i} \varepsilon_{t_i} - \frac{1}{\sqrt{N_1}} \sum_{t_i \le 1} \Delta X_{t_i} \varepsilon_{t_{i-1}}.$$

We have

$$E \left[\left(\frac{1}{\sqrt{N_1}} \sum_{t_i \le 1} \Delta X_{t_i} \varepsilon_{t_i} \right)^2 \middle| \mathcal{F}_1 \right] = \frac{\sigma_\varepsilon^2}{N_1} [X, X]_1 = O_p \left(\frac{1}{n} \right)$$

by again C(2) and $[X, X]_1 = O_p(1)$. The same argument applies to the other term. Hence, $[X, \varepsilon]_1/\sqrt{N_1} = O_p\left(1/\sqrt{n}\right)$. For the last term $([\varepsilon, \varepsilon]_1 - 2N_1\sigma_\varepsilon^2)/2\sqrt{N_1}$, we rewrite it as

$$\frac{([\varepsilon, \varepsilon]_1 - 2N_1\sigma_\varepsilon^2)}{2\sqrt{N_1}} = \frac{1}{\sqrt{N_1}}\sum_{t_i \leq 1}(\varepsilon_{t_i}^2 - \sigma_\varepsilon^2) - \frac{1}{\sqrt{N_1}}\sum_{t_i \leq 1}\varepsilon_{t_{i-1}}\varepsilon_{t_i} - \frac{\varepsilon_{t_0}^2 + \varepsilon_{t_{N_1}}^2 - 2\sigma_\varepsilon^2}{2\sqrt{N_1}}.$$

Similarly as above, we have

$$E\left[\left(\frac{1}{\sqrt{N_1}}\sum_{t_i \leq 1}(\varepsilon_{t_i}^2 - \sigma_\varepsilon^2)\right)^2 \middle| \mathcal{F}_1\right] = \left(1 + \frac{1}{N_1}\right)\text{Var}(\varepsilon^2) = O_p(1),$$

and $1/\sqrt{N_1}\sum_{t_i \leq 1}\varepsilon_{t_{i-1}}\varepsilon_{t_i} = O_p(1)$ and $(\varepsilon_{t_0}^2 + \varepsilon_{t_{N_1}}^2 - 2\sigma_\varepsilon^2)/(2\sqrt{N_1}) = O_p(1/\sqrt{n})$, completing the proof. $\square$

Next, as we will deal with sums of a random number of random variables repeatedly, the following simple lemma turns out to be very useful.

Lemma 2. *Suppose that N is a random variable taking values in nonnegative integers, and $X_1, X_2, \ldots$ are nonnegative random variables satisfying*

$$E(X_i I_{\{i \leq N\}}) \leq C \cdot P(i \leq N), \quad \text{for all } i.$$

Then

$$E\sum_{i=1}^{N} X_i \leq C \cdot E(N).$$

Proof. The conclusion follows from the fact that $\sum_{i=1}^{N} X_i = \sum_{i=1}^{\infty} X_i I_{\{i \leq N\}}$ and the Monotone Convergence theorem. $\square$

A.2. Proof of Theorem 1: single sub-grid case

The basic idea is to decompose

$$\widehat{\langle X, X \rangle}_t^{LA} - \langle X, X \rangle_t = [\overline{Y}, \overline{Y}]_t^S - \frac{2L_t}{p}\widehat{\sigma_\varepsilon^2} - \langle X, X \rangle_t$$

into existing familiar quantities and other negligible terms. The proof is divided into three steps.

Step 1: Introducing $\widetilde{Y}$

The local average can be decomposed as follows

$$\begin{aligned}
\overline{Y}_{t_{i,0}} &= \frac{1}{p}\sum_{j=0}^{p-1}(X_{t_{iq+p-j}} + \varepsilon_{t_{iq+p-j}}) \\
&= X_{t_{i,0}} - \frac{1}{p}\sum_{j=0}^{p-1}(X_{t_{iq+p}} - X_{t_{iq+p-j}}) + \frac{1}{p}\sum_{j=0}^{p-1}\varepsilon_{t_{iq+p-j}} \\
&= X_{t_{i,0}} - \sum_{j=2}^{p}\frac{j-1}{p}\Delta X_{t_{iq+j}} + \bar{\varepsilon}_{t_{i,0}}
\end{aligned}$$

where

$$\bar{\varepsilon}_{t_{i,0}} := \frac{1}{p} \sum_{j=0}^{p-1} \varepsilon_{t_{iq+p-j}},$$

which is a sequence of independent random variables with common mean $E\bar{\varepsilon} = 0$, variance $E\bar{\varepsilon}^2 = \sigma_\varepsilon^2/p$, $E\bar{\varepsilon}^3 = E\varepsilon^3/p^2$ and $E\bar{\varepsilon}^4 = E\varepsilon^4/p^3 + 3(p-1)(\sigma_\varepsilon^2)^2/p^3$. Motivated by the above decomposition, we introduce the new process $\widetilde{Y}$ as follows

$$\widetilde{Y}_{t_{i,0}} = X_{t_{i,0}} + \bar{\varepsilon}_{t_{i,0}}, \quad \text{for } i = 0, \dots, L_1.$$

The strategy is that if the difference $([\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}} - [\overline{Y}, \overline{Y}]_t^{\mathcal{S}})$, where similarly to the definition of $[\overline{Y}, \overline{Y}]_t^{\mathcal{S}}$

$$[\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}} := \sum_{t_{i,0} \leq t} (\Delta \widetilde{Y}_{t_{i,0}})^2 \quad \text{and} \quad \Delta \widetilde{Y}_{t_{i,0}} = \widetilde{Y}_{t_{i,0}} - \widetilde{Y}_{t_{i-1,0}},$$

is of a negligible order, then one needs only to deal with $[\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}}$.

Step 2: Determining the order of $([\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}} - [\overline{Y}, \overline{Y}]_t^{\mathcal{S}})$

For notational convenience, we define for $k = 0, 1, \dots, q-1$,

$$\begin{cases} A_i^k := X_{t_{i,0}^k}, \\ B_i^k := \bar{\varepsilon}_{t_{i,0}^k}, \quad \text{and} \\ C_i^k := -\sum_{j=2}^{p} \frac{j-1}{p} \Delta X_{t_{iq+j+k}}, \end{cases} \tag{A.1}$$

and let

$$A_i = A_i^0, \qquad B_i = B_i^0, \quad \text{and} \quad C_i = C_i^0.$$

Adopting the above notation, we can write

$$[\overline{Y}, \overline{Y}]_t^{\mathcal{S}} - [\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}} = \sum_{t_{i,0} \leq t} (\Delta A_i + \Delta B_i + \Delta C_i)^2 - \sum_{t_{i,0} \leq t} (\Delta A_i + \Delta B_i)^2$$

$$= \underbrace{\sum_{t_{i,0} \leq t} \Delta C_i^2}_{\text{I}} + \underbrace{2 \sum_{t_{i,0} \leq t} \Delta A_i \Delta C_i}_{\text{II}} + \underbrace{2 \sum_{t_{i,0} \leq t} \Delta B_i \Delta C_i}_{\text{III}}. \tag{A.2}$$

By the Cauchy–Schwarz inequality, for any t,

$$\text{I} \leq 4 \sum_{t_{i,0} \leq 1} C_i^2 \leq 4 \sum_{t_{iq} \leq 1} C_i^2.$$

By the BDG inequality and the strong Markov property of X,

$$E[C_i^2 I_{\{t_{iq} \leq 1\}}] = E\left(I_{\{t_{iq} \leq 1\}} E\left[C_i^2 | \mathcal{F}_{t_{iq}} \right] \right)$$

$$\leq CE\left[I_{\{t_{iq} \leq 1\}} E\left(\sum_{j=1}^{p-1} \frac{j^2}{p^2} \int_{t_{iq+j}}^{t_{iq+j+1}} \sigma_s^2 ds \, \bigg| \mathcal{F}_{t_{iq}} \right) \right]$$

$$\leq CE\left[I_{\{t_{iq}<1\}}E\left(\sum_{j=1}^{p-1}\frac{j^2}{p^2}\sigma_+^2\Delta_n\,\middle|\,\mathcal{F}_{t_{iq}}\right)\right]\leq C\frac{p}{n^{1-\eta}}P(t_{iq}\leq 1).$$

By Lemma 2 and the fact that $N_1\leq Cn$ and hence $L_1\leq Cn/q$ we then obtain

$$E(\mathrm{I})\leq 4E\left(\sum_{t_{i,0}\leq 1}C_i^2\right)\leq 4E\left(\sum_{t_{iq}\leq 1}C_i^2\right)\leq Cp/(qn^{-\eta}). \tag{A.3}$$

Next we study term III. In fact,

$$E(\mathrm{III})^2=4E\left[E\left(\left(\sum_{t_{i,0}\leq t}\Delta B_i\Delta C_i\right)^2\,\middle|\,\mathcal{F}_1\right)\right]\leq C\frac{\sigma_\varepsilon^2}{p}E\left(\sum_{t_{i,0}\leq t}C_i^2\right). \tag{A.4}$$

Hence, it follows from (A.3) that $\mathrm{III}=O_p\left((1/(qn^{-\eta}))^{1/2}\right)$.

Finally we deal with term II.

Claim 1. $\mathrm{II}=2\sum_{t_{i,0}\leq t}\Delta A_i\Delta C_i=O_p\left(\frac{\ell p}{n^{1-\eta}}\right)+O_p\left(\sqrt{\frac{p}{n^{1-2\eta}}}\right).$

Proof of Claim 1. First notice that

$$\sum_{t_{i,0}\leq t}\Delta A_i\Delta C_i=\sum_{t_{i,0}\leq t}C_i\Delta A_i-\sum_{t_{i,0}\leq t}C_{i-1}\Delta A_i,$$

where, by the BDG inequality and (A.3), we have that

$$E\left(\sum_{t_{i,0}\leq t}C_{i-1}\Delta A_i\right)^2\leq CE\sum_{t_{i,0}\leq t}C_{i-1}^2\frac{q}{n^{1-\eta}}\sigma_+^2\leq C\ell\frac{pq}{n^{2-2\eta}}\leq C\frac{p}{n^{1-2\eta}}, \tag{A.5}$$

and hence $\sum_{t_{i,0}\leq t}C_{i-1}\Delta A_i=O_p\left(\sqrt{\frac{p}{n^{1-2\eta}}}\right)$. Next define $\Delta X^{(i)}:=X_{t_{iq+1}}-X_{t_{(i-1)q+p}}$. Then

$$\sum_{t_{i,0}\leq t}C_i\Delta A_i=\underbrace{-\sum_{t_{i,0}\leq t}\sum_{j=2}^{p}\frac{j-1}{p}\left(\Delta X_{t_{iq+j}}\right)^2}_{\varsigma_1}$$

$$-\underbrace{\sum_{t_{i,0}\leq t}\sum_{j=2}^{p}\left[\frac{j-1}{p}\Delta X^{(i)}+\frac{1}{p}\sum_{m=2}^{j-1}(j+m-2)\Delta X_{t_{iq+m}}\right]\Delta X_{t_{iq+j}}}_{\varsigma_2}. \tag{A.6}$$

By the BDG inequality, $E\varsigma_1\leq C\ell p/n^{1-\eta}$; moreover, by the BDG inequality again,

$$E\left[\frac{j-1}{p}\Delta X^{(i)}+\frac{1}{p}\sum_{m=2}^{j-1}(j+m-2)\Delta X_{t_{iq+m}}\right]^2\leq Cq/n^{1-\eta};$$

hence, applying once more the BDG inequality one obtains that

$$E(\varsigma_2)^2\leq C\frac{\ell pq}{n^{2-2\eta}}. \tag{A.7}$$

It follows that $\varsigma_2 = O_p\left((p/n^{1-2\eta})^{1/2}\right)$ and moreover,

$$II = O_p\left(\frac{\ell p}{n^{1-\eta}}\right) + O_p\left(\sqrt{\frac{p}{n^{1-2\eta}}}\right). \quad \square$$

To summarize,

$$[\overline{Y}, \overline{Y}]_t^{\mathcal{S}} - [\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}} = O_p\left(\frac{\ell p}{n^{1-\eta}}\right) + O_p\left(\sqrt{\frac{p}{n^{1-2\eta}}}\right). \tag{A.8}$$

Remark 4. In the proof of Theorem 2, we will analyze $[\overline{Y}, \overline{Y}]_t^{\mathcal{S}} - [\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}}$ in more detail. Notice that $I = 2\sum_{t_{i,0}\leq t} C_i^2 - 2\sum_{t_{i,0}\leq t} C_{i-1}C_i - C_0^2 - C_{L_t}^2$, where the end effect terms C_0^2 and $C_{L_t}^2$ are $O_p\left((p/n^{1-\eta})^{1/2}\right)$ and by the BDG inequality, $\sum_{t_{i,0}\leq t} C_{i-1}C_i$ is $O_p\left(p\ell^{1/2}/n^{1-\eta}\right)$. Hence, from (A.2) and the analysis of terms I, II and III,

$$[\overline{Y}, \overline{Y}]_t^{\mathcal{S}} - [\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}} = 2\sum_{t_{i,0}\leq t}\left(\sum_{j=2}^{p} \frac{j-1}{p}\Delta X_{t_{iq+j}}\right)^2$$

$$- 2\sum_{t_{i,0}\leq t}\sum_{j=2}^{p} \frac{j-1}{p}(\Delta X_{t_{iq+j}})^2 + O_p\left(\sqrt{\frac{p}{n^{1-2\eta}}}\right). \tag{A.9}$$

Moreover,

$$\left(\sum_{j=2}^{p} \frac{j-1}{p}\Delta X_{t_{iq+j}}\right)^2 = \sum_{j=2}^{p} \frac{(j-1)^2}{p^2}(\Delta X_{t_{iq+j}})^2$$

$$+ 2\sum_{2\leq k<j\leq p} \frac{(k-1)(j-1)}{p^2}\Delta X_{t_{iq+k}}\Delta X_{t_{iq+j}}. \tag{A.10}$$

Step 3: CLT for $\widehat{\langle X, X\rangle}_t^{LA}$

We first notice that

$$[\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}} = \sum_{t_{i,0}\leq t}(\Delta X_{t_{i,0}})^2 + 2\sum_{t_{i,0}\leq t}(\Delta X_{t_{i,0}})(\Delta\bar\varepsilon_{t_{i,0}}) + \sum_{t_{i,0}\leq t}(\Delta\bar\varepsilon_{t_{i,0}})^2$$

$$:= [X, X]_t^{\mathcal{S}} + 2[X, \bar\varepsilon]_t^{\mathcal{S}} + [\bar\varepsilon, \bar\varepsilon]_t^{\mathcal{S}}. \tag{A.11}$$

Hence, we have the following decomposition

$$\widehat{\langle X, X\rangle}_t^{LA} - \langle X, X\rangle_t = [\overline{Y}, \overline{Y}]_t^{\mathcal{S}} - [\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}} + [\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}} - \langle X, X\rangle_t - \frac{2L_t}{p}\widehat{\sigma_\varepsilon^2}$$

$$= \left([\overline{Y}, \overline{Y}]_t^{\mathcal{S}} - [\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}}\right) + \left([X, X]_t^{\mathcal{S}} - \langle X, X\rangle_t\right)$$

$$+ \left([\bar\varepsilon, \bar\varepsilon]_t^{\mathcal{S}} - \frac{2L_t}{p}\sigma_\varepsilon^2\right) - \frac{2L_t}{p}\left(\widehat{\sigma_\varepsilon^2} - \sigma_\varepsilon^2\right) + 2[X, \bar\varepsilon]_t^{\mathcal{S}}. \tag{A.12}$$

Recall that $\ell \sim C_\ell n^\alpha$ and $p \sim C_p n^\alpha$. Then by (A.8), $[\overline{Y}, \overline{Y}]_t^\mathcal{S} - [\widetilde{Y}, \widetilde{Y}]_t^\mathcal{S}$ is $o_p(1/\sqrt{\ell})$. As to the term $2L_t/p(\widehat{\sigma_\varepsilon^2} - \sigma_\varepsilon^2)$ in (A.12), by Lemma 1 together with C(2) and C(4), we have that

$$\frac{2L_t}{p}\left(\widehat{\sigma_\varepsilon^2} - \sigma_\varepsilon^2\right) = \frac{2L_t}{p\sqrt{N_1}}\sqrt{N_1}\left(\widehat{\sigma_\varepsilon^2} - \sigma_\varepsilon^2\right) = O_p\left(\frac{\ell}{p\sqrt{n}}\right) = o_p\left(\frac{1}{\sqrt{\ell}}\right) \quad \text{in } D[0, 1].$$

Therefore, in order to prove the asymptotic property of $\sqrt{\ell}\left(\widehat{\langle X, X\rangle}_t^{LA} - \langle X, X\rangle_t\right)$, one only needs to prove the FCLT for the following quantity

$$\sqrt{\ell}\left([X, X]_t^\mathcal{S} - \langle X, X\rangle_t\right) + \sqrt{\ell}\left([\bar\varepsilon, \bar\varepsilon]_t^\mathcal{S} - \frac{2L_t}{p}\sigma_\varepsilon^2\right) + 2\sqrt{\ell}[X, \bar\varepsilon]_t^\mathcal{S}. \tag{A.13}$$

First, notice that

$$[\bar\varepsilon, \bar\varepsilon]_t^\mathcal{S} - \frac{2L_t}{p}\sigma_\varepsilon^2 = 2\sum_{i=1}^{L_t}\left(\bar\varepsilon_{t_{i,0}}^2 - \frac{\sigma_\varepsilon^2}{p}\right) - \bar\varepsilon_{t_{0,0}}^2 - \bar\varepsilon_{t_{L_t,0}}^2 - 2\sum_{i=1}^{L_t}\bar\varepsilon_{t_{i-1,0}}\bar\varepsilon_{t_{i,0}}.$$

Note that $\bar\varepsilon_{t_{0,0}}^2 = O_p(1/p)$; hence $\sqrt{\ell}\bar\varepsilon_{t_{0,0}}^2 = o_p(1)$, and so is $\sqrt{\ell}\bar\varepsilon_{t_{L_t,0}}^2$. Moreover,

$$[X, \bar\varepsilon]_t^\mathcal{S} = \sum_{i=1}^{L_t}(\Delta X_{t_{i,0}} - \Delta X_{t_{i+1,0}})\bar\varepsilon_{t_{i,0}} + \Delta X_{t_{L_t+1,0}}\bar\varepsilon_{t_{L_t,0}} - \Delta X_{t_{1,0}}\bar\varepsilon_{t_{0,0}}.$$

Note that $\Delta X_{t_{L_t+1,0}}\bar\varepsilon_{t_{L_t,0}} = O_p\left((1/(p\ell n^{-\eta}))^{1/2}\right)$ and so is $\Delta X_{t_{1,0}}\bar\varepsilon_{t_{0,0}}$. We are hence led to study the following martingales

$$M_t := \sqrt{\ell}\left([X, X]_t^\mathcal{S} - \langle X, X\rangle_t\right),$$

$$M_t^{(1)} := \sqrt{\ell}\sum_{i=1}^{L_t}\left(\bar\varepsilon_{t_{i,0}}^2 - \frac{\sigma_\varepsilon^2}{p}\right),$$

$$M_t^{(2)} := \sqrt{\ell}\sum_{i=1}^{L_t}\bar\varepsilon_{t_{i-1,0}}\bar\varepsilon_{t_{i,0}},$$

$$M_t^{(3)} := \sqrt{\ell}\sum_{i=1}^{L_t}(\Delta X_{t_{i,0}} - \Delta X_{t_{i+1,0}})\bar\varepsilon_{t_{i,0}}.$$

Then (A.13) can be rewritten as

$$M_t + 2M_t^{(1)} - 2M_t^{(2)} + 2M_t^{(3)} + o_p(1). \tag{A.14}$$

Simple calculation gives the corresponding predictable variation processes as follows

$$\langle M^{(1)}, M^{(1)}\rangle_t | \mathcal{F}_1 = \ell L_t \text{Var}(\bar\varepsilon^2) = \ell L_t\left(E\bar\varepsilon^4 - (E\bar\varepsilon^2)^2\right)$$

$$= \ell L_t\left(\frac{E\varepsilon^4}{p^3} + \frac{2(\sigma_\varepsilon^2)^2}{p^2} - \frac{3(\sigma_\varepsilon^2)^2}{p^3}\right) \xrightarrow{P} 2\left(\frac{C_\ell}{C_p}\sigma_\varepsilon^2\right)^2\int_0^t r_s ds,$$

$$\langle M^{(2)}, M^{(2)}\rangle_t | \mathcal{F}_1 = \ell\frac{\sigma_\varepsilon^2}{p}\sum_{i=1}^{L_t}\bar\varepsilon_{t_{i-1,0}}^2 \xrightarrow{P} \left(\frac{C_\ell}{C_p}\sigma_\varepsilon^2\right)^2\int_0^t r_s ds, \quad \text{and}$$

$$\langle M^{(3)}, M^{(3)} \rangle_t | \mathcal{F}_1 = 2\ell \frac{\sigma_\varepsilon^2}{p} [X, X]_t^S - 2\ell \frac{\sigma_\varepsilon^2}{p} \sum_{i=1}^{L_t} (\Delta X_{t_{i,0}})(\Delta X_{t_{i+1,0}})$$

$$- \ell \frac{\sigma_\varepsilon^2}{p} \left((\Delta X_{t_{1,0}})^2 + (\Delta X_{t_{L_t+1,0}})^2 \right) \xrightarrow{p} 2 \frac{C_\ell}{C_p} \langle X, X \rangle_t \sigma_\varepsilon^2,$$

where in the last convergence we used the fact that $\ell/p \cdot \sum_{i=1}^{L_t} (\Delta X_{t_{i,0}})(\Delta X_{t_{i+1,0}}) \to 0$ in $D[0, 1]$ since it is a martingale with predictable variation

$$\frac{\ell^2}{p^2} \sum_{i=1}^{L_t} (\Delta X_{t_{i,0}})^2 \int_{t_{i,0}}^{t_{i+1,0}} \sigma_s^2 \, ds \le C \frac{\ell^2 q}{p^2 n^{1-\eta}} \sum_{i=1}^{L_t} (\Delta X_{t_{i,0}})^2 = O_p \left(\frac{\ell}{p^2 n^{-2\eta}} \right) = o_p(1).$$

Furthermore, the predictable covariation processes of $M^{(1)}$, $M^{(2)}$ and $M^{(3)}$ are

$$\langle M^{(1)}, M^{(2)} \rangle_t | \mathcal{F}_1 = \ell E \bar{\varepsilon}^3 \sum_{i=1}^{L_t} \bar{\varepsilon}_{t_{i-1,0}} = \ell \frac{E \bar{\varepsilon}^3}{p^2} \sum_{i=1}^{L_t} \bar{\varepsilon}_{t_{i-1,0}} = O_p \left(\frac{\sqrt{\ell}}{p^{3/2}} \right) = o_p(1),$$

$$\langle M^{(1)}, M^{(3)} \rangle_t | \mathcal{F}_1 = \ell \sum_{i=1}^{L_t} (\Delta X_{t_{i,0}} - \Delta X_{t_{i+1,0}}) E(\bar{\varepsilon}^3)$$

$$= \ell E(\bar{\varepsilon}^3)(\Delta X_{t_{1,0}} - \Delta X_{t_{L_t+1,0}})$$

$$= O_p(\sqrt{\ell/(p^4 n^{-\eta})}) = o_p(1), \quad \text{and}$$

$$\langle M^{(2)}, M^{(3)} \rangle_t | \mathcal{F}_1 = \ell E \bar{\varepsilon}^2 \sum_{i=1}^{L_t} (\Delta X_{t_{i,0}} - \Delta X_{t_{i+1,0}}) \bar{\varepsilon}_{t_{i-1,0}} = O_p \left(\frac{\ell}{p^{3/2}} \right) = o_p(1),$$

where the last order follows from the fact that $\sum_{i=1}^{L_t} (\Delta X_{t_{i,0}} - \Delta X_{t_{i+1,0}}) \bar{\varepsilon}_{t_{i-1,0}} = O_p(1/p^{1/2})$ by considering its predictable variation process similarly to the way that we treat $M_t^{(3)}$. The Lindeberg type condition can be easily verified by using the same calculations as above and the assumption that $(\varepsilon_{t_i})_{i \ge 1}$ is an independent sequence with finite forth moment. Therefore, the usual martingale central limit theorem gives

$$\begin{pmatrix} M_t^{(1)} \\ M_t^{(2)} \\ M_t^{(3)} \end{pmatrix} \Bigg| \mathcal{F}_1 \Longrightarrow \Sigma^{1/2} \begin{pmatrix} W_1(t) \\ W_2(t) \\ W_3(t) \end{pmatrix}, \tag{A.15}$$

where W_1, W_2 and W_3 are independent standard Brownian motions and the limiting covariance matrix process is given by

$$\Sigma = \begin{pmatrix} 2 \left(\frac{C_\ell}{C_p} \sigma_\varepsilon^2 \right)^2 \int_0^t r_s ds & 0 & 0 \\ 0 & \left(\frac{C_\ell}{C_p} \sigma_\varepsilon^2 \right)^2 \int_0^t r_s ds & 0 \\ 0 & 0 & 2 \frac{C_\ell}{C_p} \langle X, X \rangle_t \sigma_\varepsilon^2 \end{pmatrix}.$$

Finally, by Theorem 1 in [20], we have the following convergence for M_t

$$M_t \Longrightarrow \frac{2}{3} \int_0^t v_s \sigma_s dX_s + \int_0^t \left(\frac{2}{3} u_s - \frac{4v_s^2}{9} \right) \sigma_s^4 \, dW(s), \tag{A.16}$$

where $W(s)$ is a standard Brownian motion. Furthermore, it is easy to see that $\langle M, M^{(i)} \rangle_t = 0$ for $i = 1, 2, 3$; hence $W(t)$ is independent of $W_i(t)$, $i = 1, 2, 3$. Combining this fact with (A.15) and (A.16) yields the desired convergence. □

A.3. Proof of Theorem 2: multiple sub-grids case

We shall establish the following stable in law convergence

$$\sqrt{\ell} \left(\frac{1}{q} \sum_{k=0}^{q-1} [\overline{Y}, \overline{Y}]_t^{\mathcal{S}_k} - \frac{2N_t}{pq} \widehat{\sigma_\varepsilon^2} - (1 + A(p,q)) \langle X, X \rangle_t \right)$$

$$\Longrightarrow \frac{2}{3} \int_0^t \bar{v}_s \sigma_s dX_s + \int_0^t \left[\left(4w_s - \frac{4}{9} \bar{v}_s^2 \right) \sigma_s^4 + \frac{8C_\ell^3}{C_p} r_s (\sigma_\varepsilon^2)^2 \right]^{1/2} dB_s.$$

Similar to the convention of using notation $[\overline{Y}, \overline{Y}]_t^{\mathcal{S}_k}$ to denote RV of local averaged Y process computed based on the kth sub-grid $\mathcal{S}_k$, all subsequent notations in the proof with superscript k or $\mathcal{S}_k$ indicate that the same operation as performed on the sub-grid $\mathcal{S} \equiv \mathcal{S}_0$ is applied to the kth sub-grid.

The proof for Theorem 2 also proceeds in three steps. Similar to the proof of Theorem 1, the proof for Theorem 2 is based on the following decomposition

$$\sqrt{\ell} \left(\frac{1}{q} \sum_{k=0}^{q-1} [\overline{Y}, \overline{Y}]_t^{\mathcal{S}_k} - \frac{2N_t}{pq} \widehat{\sigma_\varepsilon^2} - (1 + A(p,q)) \langle X, X \rangle_t \right)$$

$$= \underbrace{\sqrt{\ell} \left(\frac{1}{q} \sum_{k=0}^{q-1} [\overline{Y}, \overline{Y}]_t^{\mathcal{S}_k} - \frac{1}{q} \sum_{k=0}^{q-1} [\widetilde{Y}, \widetilde{Y}]_t^{\mathcal{S}_k} - A(p,q) \langle X, X \rangle_t \right)}_{\text{I}}$$

$$+ \underbrace{\sqrt{\ell} \left(\frac{2}{q} \sum_{k=0}^{q-1} [X, \bar{\varepsilon}]_t^{\mathcal{S}_k} + \frac{1}{q} \sum_{k=0}^{q-1} [\bar{\varepsilon}, \bar{\varepsilon}]_t^{\mathcal{S}_k} - \frac{2N_t}{pq} \widehat{\sigma_\varepsilon^2} \right)}_{\text{II}}$$

$$+ \underbrace{\sqrt{\ell} \left(\frac{1}{q} \sum_{k=0}^{q-1} [X, X]_t^{\mathcal{S}_k} - \langle X, X \rangle_t \right)}_{\text{III}}.$$

Assuming $\ell \sim C_\ell n^\alpha$ and $p \sim C_p n^{3\alpha - 1}$ with assumptions made in the theorem on α and η, we shall show in Step 1 that I $= o_p(1)$; in Step 2 that II satisfies a martingale CLT; in Step 3 a CLT with asymptotic bias decomposition for term III; and, finally, sum up in Step 4.

Step 1

To show $\mathrm{I} = o_p(1)$, we consider the difference

$$\frac{\sqrt{\ell}}{q} \sum_{k=0}^{q-1} [\overline{Y}, \overline{Y}]_t^{S_k} - \frac{\sqrt{\ell}}{q} \sum_{k=0}^{q-1} [\widetilde{Y}, \widetilde{Y}]_t^{S_k}$$

$$= \frac{\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \left(\sum_{t_{i,0}^k \le t} (\Delta C_i^k)^2 + 2 \sum_{t_{i,0}^k \le t} \Delta A_i^k \Delta C_i^k + 2 \sum_{t_{i,0}^k \le t} \Delta B_i^k \Delta C_i^k \right), \tag{A.17}$$

adopting the previous notational convention for the single sub-grid case, where A_i^k, B_i^k and C_i^k are defined in (A.1). Roughly speaking, recall (A.9) and (A.10) of Remark 4 from the end of Step 2 in the proof for Theorem 1, we expect the difference (A.17) to be

$$\frac{2\sqrt{\ell}}{q} \sum_{j=1}^{p-1} \left(\frac{j^2}{p^2} - \frac{j}{p} \right) \cdot \sum_{t_i \le t} (\Delta X_{t_i})^2 + o_p(1) = \sqrt{\ell} A(p, q)[X, X]_t + o_p(1). \tag{A.18}$$

It is easy to see that $\sqrt{\ell} A(p, q)([X, X]_t - \langle X, X \rangle_t) = o_p(1)$. Hence $I = o_p(1)$ if we can show that (A.18) holds.

We now verify (A.18). It is easy to see that the RHS of (A.17) equals

$$\underbrace{\frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \sum_{t_{i,0}^k \le t} (C_i^k)^2}_{\mathrm{I.i}} + \underbrace{\frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \sum_{t_{i,0}^k \le t} C_{i-1}^k C_i^k}_{\mathrm{I.ii}} + \underbrace{\frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \sum_{t_{i,0}^k \le t} \Delta B_i^k \Delta C_i^k}_{\mathrm{I.iii}};$$

$$\underbrace{- \frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \sum_{t_{i,0}^k \le t} C_{i-1}^k \Delta A_i^k}_{\mathrm{I.iv}} + \underbrace{\frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \sum_{t_{i,0}^k \le t} C_i^k \Delta A_i^k}_{\mathrm{I.v}} + o_p(1).$$

We analyze them one by one.

We start with $\mathrm{I.i} = \frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \sum_{t_{i,0}^k \le t} (C_i^k)^2$. Notice that on each sub-grid S_k,

$$(C_i^k)^2 = \sum_{j=1}^{p-1} \frac{j^2}{p^2} (\Delta X_{t_{iq+j+k+1}})^2 + 2 \sum_{j=2}^{p-1} \left(\sum_{m=1}^{j-1} \frac{m}{p} \Delta X_{t_{iq+m+k+1}} \right) \frac{j}{p} \Delta X_{t_{iq+j+k+1}}.$$

Therefore, term I.i can be rewritten as follows

$$\underbrace{\frac{2\sqrt{\ell}}{q} \left(\sum_{j=1}^{p-1} \frac{j^2}{p^2} \right) \sum_{t_i \le t} (\Delta X_{t_i})^2 - \frac{2\sqrt{\ell}}{q} \left(\sum_{i=1}^{p-1} \left(\sum_{j=i}^{p-1} \frac{j^2}{p^2} \right) (\Delta X_{t_i})^2 \right)}_{\text{dominating term } A}$$

$$+ \sum_{i=L_t q+2}^{L_t q+p-1} \left(\sum_{j=1}^{i-L_t q-1} \frac{j^2}{p^2} \right) (\Delta X_{t_{i+1}})^2 \Bigg)$$

$$
\begin{aligned}
&+ \left(\frac{4\sqrt{\ell}}{q} \sum_{i=2}^{p-2} \left(\sum_{j=2}^{i} \frac{j}{p} \sum_{m=1}^{j-1} \frac{m}{p} \Delta X_{t_{i-j+m+1}} \right) \Delta X_{t_{i+1}} \right. \\
&+ \frac{4\sqrt{\ell}}{q} \sum_{i=p-1}^{L_t q+2} \left(\sum_{j=2}^{p-1} \frac{j}{p} \sum_{m=1}^{j-1} \frac{m}{p} \Delta X_{t_{i-j+m+1}} \right) \Delta X_{t_{i+1}} \\
&+ \left. \frac{4\sqrt{\ell}}{q} \sum_{i=L_t q+3}^{L_t q+p-1} \left(\sum_{j=i-L_t q}^{p-1} \frac{j}{p} \sum_{m=1}^{j-1} \frac{m}{p} \Delta X_{t_{i-j+m+1}} \right) \Delta X_{t_{i+1}} \right)
\end{aligned}
$$

$$
:= \text{ dominating term } A - \text{ edge term } B + S_t^{(1)}. \tag{A.19}
$$

It is easy to see that the edge term $B = o_p(1)$. We shall further show that $S_t^{(1)}$ is negligible. To see that, notice that its expected predictable variation satisfies

$$
\begin{aligned}
E\langle S^{(1)}, S^{(1)} \rangle_1 &\leq \frac{C\ell\sigma_+^2}{q^2 n^{1-\eta}} E \sum_i \left(\sum_{j=2}^{p-1} \frac{j}{p} \sum_{m=1}^{j-1} \frac{m}{p} \Delta X_{t_{i-j+m+1}} \right)^2 \\
&= \frac{C\ell\sigma_+^2}{q^2 n^{1-\eta}} E \sum_i \left(\sum_{m=1}^{p-2} \left(\sum_{j=m+1}^{p-1} \frac{j}{p} \frac{j-m}{p} \right) \Delta X_{t_{i-m+1}} \right)^2 \\
&\leq \frac{C\ell p^3}{n^{1-2\eta} q^2},
\end{aligned}
$$

which follows from the fact that

$$
E \left(\sum_{m=1}^{p-2} \left(\sum_{j=m+1}^{p-1} \frac{j}{p} \frac{j-m}{p} \right) \Delta X_{t_{i-m+1}} \right)^2 \leq C \frac{p^3}{n^{1-\eta}}, \quad \text{uniformly in } i.
$$

Therefore,

$$
S_t^{(1)} = O_p \left(\sqrt{\frac{\ell p^3}{n^{1-2\eta} q^2}} \right) = O_p \left(\sqrt{\frac{p^3}{q^3 n^{-2\eta}}} \right) = o_p(1) \quad \text{in } D[0,1].
$$

Next we estimate $\text{I.ii} = \frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \sum_{t_{i,0} \leq t} C_{i-1}^k C_i^k$. It can be rearranged as

$$
\begin{aligned}
\text{I.ii} = &\frac{2\sqrt{\ell}}{q} \sum_{i=q+1}^{q+p-2} \left(\sum_{j=1}^{i-q} \frac{j}{p} \sum_{m=1}^{p-1} \frac{m}{p} \Delta X_{t_{i-q-j+m+1}} \right) \Delta X_{t_{i+1}} \\
&+ \frac{2\sqrt{\ell}}{q} \sum_{i=q+p-1}^{L_t q+1} \left(\sum_{j=1}^{p-1} \frac{j}{p} \sum_{m=1}^{p-1} \frac{m}{p} \Delta X_{t_{i-q-j+m+1}} \right) \Delta X_{t_{i+1}} \\
&+ \frac{2\sqrt{\ell}}{q} \sum_{i=L_t q+2}^{L_t q+p-1} \left(\sum_{j=i-L_t q}^{p-1} \frac{j}{p} \sum_{m=1}^{p-1} \frac{m}{p} \Delta X_{t_{i-q-j+m+1}} \right) \Delta X_{t_{i+1}}.
\end{aligned}
$$

We denote the above quantity as $S_t^{(2)}$. Similar to the treatment for I.i,

$$E\langle S^{(2)}, S^{(2)}\rangle_1 \leq \frac{C\ell\sigma_+^2}{q^2 n^{1-\eta}} E \sum_i \left(\sum_{j=1}^{p-1}\frac{j}{p}\sum_{m=1}^{p-1}\frac{m}{p}\Delta X_{t_{i-q-j+m+1}}\right)^2 \leq \frac{C}{q^2}\frac{\ell p^3}{n^{1-2\eta}}.$$

Therefore,

$$S_t^{(2)} = O_p\left(\sqrt{\frac{\ell p^3}{n^{1-2\eta}q^2}}\right) = O_p\left(\sqrt{\frac{p^3}{q^3 n^{-2\eta}}}\right) = o_p(1) \quad \text{in } D[0,1].$$

Now we study $\text{I.iii} = \frac{2\sqrt{\ell}}{q}\sum_{k=0}^{q-1}\sum_{t_{i,0}^k \leq t}\Delta B_i^k \Delta C_i^k$. Noticing that the estimate in (A.4) holds uniformly for sub-grids $\mathcal{S}_k$, hence by the Cauchy–Schwarz inequality we obtain that

$$\text{I.iii} = O_p\left(\left(\ell/(qn^{-\eta})\right)^{1/2}\right) = o_p(1) \quad \text{in } D[0,1].$$

Now we come to $\text{I.iv} = \frac{2\sqrt{\ell}}{q}\sum_{k=0}^{q-1}\sum_{t_{i,0}^k \leq t}C_{i-1}^k\Delta A_i^k$. By the Cauchy–Schwarz inequality again, as the estimate in (A.5) holds uniformly for sub-grids $\mathcal{S}_k$, we have

$$\text{I.iv} = O_p\left(\left(\ell p/n^{1-2\eta}\right)^{1/2}\right) = o_p(1) \quad \text{in } D[0,1].$$

Finally we deal with $\text{I.v} = \frac{2\sqrt{\ell}}{q}\sum_{k=0}^{q-1}\sum_{t_{i,0}^k \leq t}C_i^k\Delta A_i^k$. Similar to the decomposition (A.6) we have

$$\text{I.v} = -\frac{2\sqrt{\ell}}{q}\sum_{k=0}^{q-1}(\varsigma_1^k + \varsigma_2^k)$$

where, with $\Delta X^{(k,i)} := X_{t_{iq+k+1}} - X_{t_{i-1q+k+p}}$,

$$\varsigma_1^k = \sum_{t_{i,0}^k \leq t}\sum_{j=2}^{p}\frac{j-1}{p}\left(\Delta X_{t_{iq+k+j}}\right)^2,$$

$$\varsigma_2^k = \sum_{t_{i,0}^k \leq t}\sum_{j=2}^{p}\left[\frac{j-1}{p}\Delta X^{(k,i)} + \frac{1}{p}\sum_{m=2}^{j-1}(j+m-2)\Delta X_{t_{iq+k+m}}\right]\Delta X_{t_{iq+k+j}}.$$

It is easy to see that

$$-2\sqrt{\ell}/q\sum_{k=0}^{q-1}\varsigma_1^k = \underbrace{-2\sqrt{\ell}/q\sum_{j=1}^{p-1}\frac{j}{p}\sum_{t_i \leq t}(\Delta X_{t_i})^2}_{\text{dominating term } B} + o_p(1).$$

We next prove that $2\sqrt{\ell}/q\sum_{k=0}^{q-1}\varsigma_2^k$ is negligible. In fact, the estimate in (A.7) holds uniformly for all the sub-grids, hence by the Cauchy–Schwarz inequality again we get that

$$2\sqrt{\ell}/q\sum_{k=0}^{q-1}\varsigma_2^k = O_p\left(\left(\ell p/n^{1-2\eta}\right)^{1/2}\right) = o_p(1) \quad \text{in } D[0,1].$$

Summing up the computations for I.i–I.v, we see that the two dominating terms appearing in I.i and I.v together give the first term in (A.18) and the rest gives the $o_p(1)$ term in (A.18).

Step 2

Now we deal with the term II, starting with $2\sqrt{\ell}/q \sum_{k=0}^{q-1} [X, \bar{\varepsilon}]_t^{S_k}$. Denote

$$\Delta_q X_{t_i} = X_{t_i} - X_{t_{i-q}}.$$

Combining terms with common factor ε_{t_i} and ordering them chronologically (according to the sequence $(\varepsilon_{t_i})_{i \geq 1}$) we get

$$\frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} [X, \bar{\varepsilon}]_t^{S_k} = \frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \left[\sum_{i=1}^{L_t^k-1} \left(\Delta X_{t_{i,0}^k} - \Delta X_{t_{i+1,0}^k} \right) \bar{\varepsilon}_{t_{i,0}^k} \right.$$

$$\left. + \Delta X_{t_{L_t^k,0}^k} \bar{\varepsilon}_{t_{L_t^k,0}^k} - \Delta X_{t_{1,0}^k} \bar{\varepsilon}_{t_{0,0}^k} \right]$$

$$= \frac{2\sqrt{\ell}}{qp} \sum_{i=q+p}^{(L_t-1)q+1} \left(\sum_{j=0}^{p-1} \left[\Delta_q X_{t_{i+j}} - \Delta_q X_{t_{i+q+j}} \right] \right) \varepsilon_{t_i} + \text{remainder},$$

where the remainder term is a sum similar as above over the i's smaller than $q + p$, and can be easily shown to be $o_p(1)$. We shall further show that the first summand is also negligible, as follows

$$\text{Var} \left(\frac{2\sqrt{\ell}}{qp} \sum_{i=q+p}^{(L_t-1)q+1} \left(\sum_{j=0}^{p-1} \left[\Delta_q X_{t_{i+j}} - \Delta_q X_{t_{i+q+j}} \right] \right) \varepsilon_{t_i} \middle| \mathcal{F}_1 \right)$$

$$= \frac{4\ell\sigma_\varepsilon^2}{q^2 p^2} \sum_i \left(\sum_{j=0}^{p-1} \left[\Delta_q X_{t_{i+j}} - \Delta_q X_{t_{i+q+j}} \right] \right)^2$$

$$= \frac{4\ell\sigma_\varepsilon^2}{q^2 p^2} \sum_i \sum_{j=0}^{p-1} \left((\Delta_q X_{t_{i+j}})^2 + (\Delta_q X_{t_{i+q+j}})^2 \right)$$

$$- \frac{8\ell\sigma_\varepsilon^2}{q^2 p^2} \sum_i \sum_{j=0}^{p-1} \Delta_q X_{t_{i+j}} \Delta_q X_{t_{i+q+j}}$$

$$+ \frac{8\ell\sigma_\varepsilon^2}{q^2 p^2} \sum_i \sum_{0 \leq j < k \leq p-1} \left[\Delta_q X_{t_{i+j}} - \Delta_q X_{t_{i+q+j}} \right] \left[\Delta_q X_{t_{i+j}} - \Delta_q X_{t_{i+q+k}} \right] \quad \text{(A.20)}$$

$$:= V_1 + V_2 + V_3. \quad \text{(A.21)}$$

We have, first, by applying Lemma 2 and using the fact that $E(\Delta_q X_{t_i})^2 \leq Cq/n^{1-\eta}$ for all i,

$$EV_1 \leq \frac{C\ell}{q^2 p^2} \cdot n \cdot p \frac{q}{n^{1-\eta}} = \frac{C\ell}{qpn^{-\eta}} \to 0.$$

This, together with the Cauchy–Schwarz inequality, imply that $E|V_2| \leq C\ell/(pqn^{-\eta}) \to 0$. Finally, using the Cauchy–Schwarz inequality again we have

$$E|V_3| \leq CpE(V_1) \leq C\ell/(qn^{-\eta}) \to 0.$$

Second, for $\sqrt{\ell}/q \sum_{k=0}^{q-1}[\bar{\varepsilon}, \bar{\varepsilon}]_t^{\mathcal{S}_k}$, following the way the terms of $2\sqrt{\ell}/q \sum_{k=0}^{q-1}[X, \bar{\varepsilon}]_t^{\mathcal{S}_k}$ were rearranged, we have

$$
\begin{aligned}
\frac{\sqrt{\ell}}{q} \sum_{k=0}^{q-1}[\bar{\varepsilon}, \bar{\varepsilon}]_t^{\mathcal{S}_k} &= \frac{\sqrt{\ell}}{pq} \sum_{t_i \le t} \left(-2 \sum_{j=0}^{p-1} \frac{p-j}{p} \varepsilon_{t_{i-q-j}} - 2 \sum_{j=1}^{p-1} \frac{p-j}{p} \varepsilon_{t_{i-q+j}} \right. \\
&\quad \left. + 4 \sum_{j=1}^{p-1} \frac{p-j}{p} \varepsilon_{t_{i-j}} \right) \varepsilon_{t_i} + \frac{2\sqrt{\ell}}{pq} \sum_{t_i \le t} \varepsilon_{t_i}^2 + o_p(1) \\
&:= M_t^{(4)} + \frac{2\sqrt{\ell}}{pq} \sum_{t_i \le t} \varepsilon_{t_i}^2 + o_p(1)
\end{aligned}
$$

where the $o_p(1)$ term is again due to the end effect.

We first deal with $M_t^{(4)}$. We need the following notation

$$
J_1 := \{1, 2, \ldots, p-1\}, \qquad J_2 := \{q-p+1, q-p+2, \ldots, q-1\} \quad \text{and}
$$
$$
J_3 := \{q, q+1, \ldots, q+p-1\}.
$$

Let $J := \bigcup_{m=1}^{3} J_m$ and $J_{\max}$ be the largest element in J. Moreover, denote the following weight function

$$
w(j) = \begin{cases} 4\dfrac{p-j}{p} & \text{for } j \in J_1; \\ -2\dfrac{p-q+j}{p} & \text{for } j \in J_2; \quad \text{and} \\ -2\dfrac{p+q-j}{p} & \text{for } j \in J_3. \end{cases}
$$

Notice that $|w(j)| \le 4$ for all $j \in J$. $M_t^{(4)}$ is a martingale with quadratic variation that can then be represented as

$$
\begin{aligned}
\langle M^{(4)}, M^{(4)} \rangle_t &= \frac{\ell \sigma_\varepsilon^2}{p^2 q^2} \sum_{t_i \le t} \left[-2 \sum_{j=0}^{p-1} \frac{p-j}{p} \varepsilon_{t_{i-q-j}} - 2 \sum_{j=1}^{p-1} \frac{p-j}{p} \varepsilon_{t_{i-q+j}} \right. \\
&\quad \left. + 4 \sum_{j=1}^{p-1} \frac{p-j}{p} \varepsilon_{t_{i-j}} \right]^2 \\
&= \frac{24 \ell \sigma_\varepsilon^2}{p^4 q^2} \sum_{t_i \le t} \sum_{j=1}^{p-1} (p-j)^2 \sigma_\varepsilon^2 + \frac{\ell \sigma_\varepsilon^2}{p^2 q^2} \sum_{t_i \le t} \sum_{j \in J} w(j)^2 (\varepsilon_{t_{i-j}}^2 - \sigma_\varepsilon^2) \\
&\quad + \frac{\ell \sigma_\varepsilon^2}{p^2 q^2} \sum_{t_i \le t} \sum_{j,k \in J, j \ne k} w(j)w(k) \varepsilon_{t_{i-j}} \varepsilon_{t_{i-k}} \\
&= \frac{24 \ell \sigma_\varepsilon^2}{p^4 q^2} \sum_{t_i \le t} \sum_{j=1}^{p-1} (p-j)^2 \sigma_\varepsilon^2 + o_p(1) \\
&\sim \frac{8n\ell(\sigma_\varepsilon^2)^2}{pq^2} \frac{N_t}{n} + o_p(1) \xrightarrow{p} \frac{8(\sigma_\varepsilon^2)^2 C_\ell^3}{C_p} \int_0^t r_s \, ds
\end{aligned}
$$

where the last line follows from the assumption that $L_t/\ell \xrightarrow{P} \int_0^t r_s ds$ and the third equality is explained as follows. We take the third term on the RHS of the second equality for example while the second term can be treated more easily by a similar argument. Notice that this term can be rewritten as

$$\mathcal{E}_t := \frac{2\ell\sigma_\varepsilon^2}{p^2 q^2} \sum_{t_i < t^*} \left[\sum_{j=1}^{J_{\max}-1} \left(\sum_{k=1}^{J_{\max}-j} w(k)w(j+k) I_{\{k \in J, \ j+k \in J\}} \right) \varepsilon_{t_i-j} \right] \varepsilon_{t_i},$$

where $t^* := \max\{t_j \leq t\}$. Hence by Lemma 2, the BDG inequality, the boundedness of $w(\cdot)$ function and the fact that the cardinality of set J is of order p, we have

$$E\left(\mathcal{E}_t^2\right) \leq \frac{4\ell^2 \sigma_\varepsilon^6}{p^4 q^4} E \sum_{t_i < t^*} \left[\sum_{j=1}^{J_{\max}-1} \left(\sum_{k=1}^{J_{\max}-j} w(k)w(j+k) I_{\{k \in J, \ j+k \in J\}} \right) \varepsilon_{t_i-j} \right]^2$$

$$\leq C \frac{nqp^2 \ell^2 \sigma_\varepsilon^8}{p^4 q^4} = C \frac{n\ell^2}{p^2 q^3} = o(1).$$

Hence $\mathcal{E}_t = o_p(1)$.

Therefore, based on our moment assumption for $(\varepsilon_{t_i})_{i \geq 1}$, $M^{(4)}$ satisfies a CLT where the limiting distribution is a mixture of normal and the mixture component is the variance equal to $\frac{8(\sigma_\varepsilon^2)^2 C_\ell^3}{C_p} \int_0^t r_s ds$; in other words,

$$M_t^{(4)} \Longrightarrow \int_0^t \left[\frac{8(\sigma_\varepsilon^2)^2 C_\ell^3}{C_p} r_s \right]^{1/2} dB_s, \tag{A.22}$$

where B_t is a standard Brownian motion that is independent of $\mathcal{F}_1$.

As to $2\sqrt{\ell}/(pq) \sum_i \varepsilon_{t_i}^2$, it follows from Lemma 1 and C(4) that

$$\frac{2\sqrt{\ell}}{pq} \sum_{t_i \leq t} \varepsilon_{t_i}^2 - \frac{2\sqrt{\ell}}{pq} N_t \widehat{\sigma_\varepsilon^2} = \frac{2\sqrt{\ell}}{pq} \sum_{t_i \leq t} (\varepsilon_{t_i}^2 - \sigma_\varepsilon^2) - \frac{2\sqrt{\ell}}{pq} N_t \left(\widehat{\sigma_\varepsilon^2} - \sigma_\varepsilon^2 \right)$$

$$= O_p\left(\frac{\sqrt{\ell n}}{pq} \right) = o_p(1) \quad \text{in } D[0, 1].$$

Step 3

Finally, we prove a CLT for term III. We have

$$\overline{M}_t := \sqrt{\ell} \left(\frac{1}{q} \sum_{k=0}^{q-1} [X, X]_t^{S_k} - \langle X, X \rangle_t \right)$$

$$= \frac{1}{q} \sum_{k=0}^{q-1} \sqrt{\ell} \left([X, X]_t^{S_k} - \langle X, X \rangle_t \right) = \frac{1}{q} \sum_{k=0}^{q-1} M_t^k,$$

where

$$dM_t^k = 2\sqrt{\ell}(X_t - X_{t_*^k}) dX_t$$

and t_*^k is the largest time smaller than or equal to t on the kth sub-grid. Therefore

$$\overline{M}_t = \frac{1}{q} \sum_{k=0}^{q-1} \int_0^t dM_s^k = \frac{2\sqrt{\ell}}{q} \sum_{k=0}^{q-1} \int_0^t (X_s - X_{t_*^k}) dX_s = 2\sqrt{\ell} \int_0^t f_n(s) dX_s,$$

where

$$f_n(s) = \begin{cases} 0, & \text{for } s \in [0, t_p); \\ \frac{1}{q} \sum_{j=0}^{i-p} (X_s - X_{t_{i-j}}), & \text{for } s \in [t_i, t_{i+1}) \text{ and } p \le i < q + p; \\ X_s - X_{t_i} + \sum_{j=1}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j+1}}, & \text{for } s \in [t_i, t_{i+1}) \text{ and } i \ge q + p. \end{cases}$$

$\overline{M}_t$ is a martingale with quadratic variation $\langle \overline{M}, \overline{M} \rangle_t = 4\ell \int_0^t f_n(s)^2 \sigma_s^2 ds$. Since $Ef_n(s)^2 \le Cq/n^{1-\eta}$,

$$4\ell \int_0^{t_{q+p}} f_n(s)^2 \sigma_s^2 ds = o_p(1). \tag{A.23}$$

Hence, we need to only consider $s \ge t_{q+p}$, i.e., $i \ge q + p$. By Itô's formula,

$$df_n(s)^4 = 4 f_n(s)^3 dX_s + 6 f_n(s)^2 \sigma_s^2 ds, \quad \text{for } s \in [t_i, t_{i+1}) \text{ and } i \ge q + p.$$

Hence

$$\langle \overline{M}, \overline{M} \rangle_t = 4\ell \int_0^t f_n(s)^2 \sigma_s^2 ds = \frac{2\ell}{3} \int_0^t df_n(s)^4 - \frac{8\ell}{3} \int_0^t f_n(s)^3 dX_s. \tag{A.24}$$

We first prove that the second term on the RHS of (A.24) is negligible. In fact, by the BDG inequality,

$$Ef_n(s)^6 \le C(q/n^{1-\eta})^3 \tag{A.25}$$

uniformly in s. Hence

$$\begin{aligned} E\left(\int_0^t f_n(s)^3 dX_s\right)^2 &\le E\left\langle \int_0^\cdot f_n(s)^3 dX_s, \int_0^\cdot f_n(s)^3 dX_s \right\rangle_t \\ &\le E \int_0^1 f_n(s)^6 \sigma_s^2 ds \\ &\le \sigma_+^2 \int_0^1 Ef_n(s)^6 ds \\ &\le C \int_0^1 \left(\frac{q}{n^{1-\eta}}\right)^3 ds = O\left(\frac{q^3}{n^{3-3\eta}}\right), \end{aligned} \tag{A.26}$$

and $\ell \int_0^t f_n(s)^3 dX_s \xrightarrow{P} 0$, in $D[0, 1]$, as $n \to \infty$. Now we deal with the first term in (A.24). We shall only focus on the integral on $[t_{i_{q+p}}, t_*]$ where t_* is the largest $t_i \le t$; the remainder term

is negligible. We then have

$$
\ell \int_{t_{i_q+p}}^{t_*} df_n(s)^4 = \ell \sum_{t_i \le t} \left\{ \left(\sum_{j=0}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j}} \right)^4 - \left(\sum_{j=1}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j}} \right)^4 \right\}
$$

$$
= \ell \underbrace{\sum_{t_i \le t} (\Delta X_{t_i})^4}_{\widetilde{\mathrm{I}}} + \ell \underbrace{\sum_{t_i \le t} 6 (\Delta X_{t_i})^2 \left(\sum_{j=1}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j}} \right)^2}_{\widetilde{\mathrm{II}}}
$$

$$
+ \ell \underbrace{\sum_{t_i \le t} 4 (\Delta X_{t_i})^3 \left(\sum_{j=1}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j}} \right)}_{\widetilde{\mathrm{III}}}
$$

$$
+ \ell \underbrace{\sum_{t_i \le t} 4 \Delta X_{t_i} \left(\sum_{j=1}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j}} \right)^3}_{\widetilde{\mathrm{IV}}}.
$$

By the BDG inequality and Lemma 2, $\widetilde{\mathrm{I}} = O_p(\ell/n^{1-2\eta}) = o_p(1)$. Moreover, for term $\widetilde{\mathrm{IV}}$, by computing its quadratic variation and using (A.25) we get

$$
\widetilde{\mathrm{IV}} = O_p(\ell q^{3/2}/n^{3/2-2\eta}) = O_p(1/(\sqrt{\ell} n^{-2\eta})) = o_p(1).
$$

As to term $\widetilde{\mathrm{III}}$, we treat it in the same fashion as above by defining for $s \in [t_{i-1}, t_i)$,

$$
\check{X}_s := 4\ell (X_s - X_{t_{i-1}})^3 \left(\sum_{j=1}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j}} \right);
$$

$$
\check{X}_s^{(1)} := 12\ell (X_s - X_{t_{i-1}})^2 \left(\sum_{j=1}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j}} \right);
$$

$$
\check{X}_s^{(2)} := 12\ell (X_s - X_{t_{i-1}}) \left(\sum_{j=1}^{q-1} \frac{q-j}{q} \Delta X_{t_{i-j}} \right).
$$

Then

$$
\widetilde{\mathrm{III}} = \sum_{t_i \le t} \int_{t_{i-1}}^{t_i} d\check{X}_s = \int_0^t \check{X}_s^{(1)} dX_s + \int_0^t \check{X}_s^{(2)} \sigma_s^2 ds + o_p(1),
$$

which is again an $o_p(1)$ term by noting that (1) $E(\check{X}_s^{(1)})^2 \le C\ell^2 \cdot 1/n^{2-2\eta} \cdot q/n^{1-\eta} \le C\ell/n^{2-3\eta}$; and (2) $E(\check{X}_s^{(2)})^2 \le C\ell^2 \cdot 1/n^{1-\eta} \cdot q/n^{1-\eta} \le C\ell/n^{1-2\eta}$. Finally, by assumption C(6) we get the convergence of term $\widetilde{\mathrm{II}}$ and hence

$$
\langle \overline{M}, \overline{M} \rangle_t \xrightarrow{p} 4 \int_0^t w_s \sigma_s^4 \, ds \quad \text{for all } t.
$$

Next, we estimate the quadratic covariation between $\overline{M}$ and X. To do so, we first notice that, by Itô's formula,

$$d\langle X, \overline{M}\rangle_t = \frac{1}{q}\sum_{k=0}^{q-1} d\langle X, M^k\rangle_t = \frac{1}{q}\sum_{k=0}^{q-1} \frac{2\sqrt{\ell}}{3} d(X_t - X_{t_*^k})^3$$
$$- \frac{1}{q}\sum_{k=0}^{q-1} 2\sqrt{\ell}(X_t - X_{t_*^k})^2 dX_t, \tag{A.27}$$

where

$$\frac{1}{q}\sum_{k=0}^{q-1} \frac{2\sqrt{\ell}}{3} d(X_t - X_{t_*^k})^3 = \frac{2}{3}\frac{1}{q}\sum_{k=0}^{q-1} \sqrt{\ell}\, d[X, X, X]_t^{\mathcal{S}_k}.$$

We next show that the martingale term in (A.27) is negligible. Rearranging terms in the same way as we did for $\overline{M}_t$, we have

$$R_t := \int_0^t \frac{1}{q}\sum_{k=0}^{q-1} 2\sqrt{\ell}(X_s - X_{t_*^k})^2 dX_s = \frac{2\sqrt{\ell}}{q}\int_0^t g_n(s)dX_s,$$

where

$$g_n(s) = \begin{cases} 0, & \text{for } s \in [0, t_p); \\ \displaystyle\sum_{j=0}^{i-p}(X_s - X_{t_{i-j}})^2, & \text{for } s \in [t_i, t_{i+1}) \text{ and } p \le i < q + p; \\ \displaystyle\sum_{j=0}^{q-1}(X_s - X_{t_{i-j}})^2, & \text{for } s \in [t_i, t_{i+1}) \text{ and } i \ge q + p. \end{cases}$$

Observe that by the Cauchy–Schwarz inequality and the BDG inequality,

$$E\left[\sum_{j=0}^{q-1}(X_{t_{i+1}} - X_{t_{i-j}})^2\right]^2 = E\sum_{j=0}^{q-1}(X_{t_{i+1}} - X_{t_{i-j}})^4 + 2E$$
$$\times \sum_{0 \le j < k \le q-1}(X_{t_{i+1}} - X_{t_{i-j}})^2(X_{t_{i+1}} - X_{t_{i-k}})^2$$
$$\le C\sigma_+^4 q\frac{q^2}{n^{2-2\eta}} + C\sigma_+^4 q(q-1)\frac{q^2}{n^{2-2\eta}} = O\left(\frac{q^4}{n^{2-2\eta}}\right).$$

Hence, uniformly in $s \in [t_i, t_{i+1})$ and i, $Eg_n(s)^2 \le Cq^4/n^{2-2\eta}$. Therefore, by the BDG inequality again,

$$E(R_t)^2 \le C\frac{\ell}{q^2}E\int_0^t g_n(s)^2\sigma_s^2 ds \le C\frac{\ell q^4}{n^{2-2\eta}q^2} \le C\frac{1}{\ell n^{-2\eta}} \to 0,$$

and hence $R_t = o_p(1)$. Therefore, assumption C(7) and (A.27) imply that

$$\langle X, \overline{M}\rangle_t \xrightarrow{p} \frac{2}{3}\int_0^t \bar{v}_s\sigma_s^3 ds \quad \text{for all } t.$$

It follows from the limit results in either Theorem B.4 (pp. 65–67) of Zhang [32] or Theorem 2.28 of Mykland and Zhang [25] that, stably in law,

$$\overline{M}_t \Longrightarrow \frac{2}{3} \int_0^t \bar{v}_s \sigma_s dX_s + \int_0^t \left[\left(4w_s - \frac{4}{9} \bar{v}_s^2 \right) \sigma_s^4 \right]^{1/2} dB_s, \tag{A.28}$$

where B_t is a standard Brownian motion that is independent of $\mathcal{F}_1$.

Step 4

Clearly $\langle M^4, \overline{M} \rangle_t = 0$. The overall results then follows from (A.22) and (A.28). □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.spa.2013.04.002.

References

[1] Y. Aït-Sahalia, J. Fan, Y. Li, The leverage effect puzzle: disentangling sources of bias at high frequency, Journal of Financial Economics (2012) (in press).
[2] Y. Aït-Sahalia, R. Kimmel, Maximum likelihood estimation of stochastic volatility models, Journal of Financial Economics 83 (2007) 413–452.
[3] Y. Aït-Sahalia, P.A. Mykland, The effects of random and discrete sampling when estimating continuous time diffusions, Econometrica 71 (2003) 483–549.
[4] Y. Aït-Sahalia, P.A. Mykland, L. Zhang, How often to sample a continuous-time process in the presence of market microstructure noise, Review of Financial Studies 18 (2005) 351–416.
[5] F.M. Bandi, J.R. Russell, Separating microstructure noise from volatility, Journal of Financial Economics 79 (2006) 655–692.
[6] O.E. Barndorff-Nielsen, P.R. Hansen, A. Lunde, N. Shephard, Designing realized Kernels to measure the ex post variation of equity prices in the presence of noise, Econometrica 76 (2008) 1481–1536.
[7] O.E. Barndorff-Nielsen, N. Shephard, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, Journal of the Royal Statistical Society: Series B 64 (2002) 253–280.
[8] D. Duffie, P. Glynn, Estimation of continuous-time Markov processes sampled at random times, Econometrica 72 (2004) 1773–1808.
[9] J. Fan, Y. Wang, Multi-scale jump and volatility analysis for high-frequency financial data, Journal of the American Statistical Association 102 (2007) 1349–1362.
[10] M. Fukasawa, Central limit theorem for the realized volatility based on tick time sampling, Finance and Stochastics 14 (2010) 209–233.
[11] M. Fukasawa, Realized volatility with stochastic sampling, Stochastic Processes and their Applications 120 (2010) 829–852.
[12] M. Fukasawa, M. Rosenbaum, Central limit theorems for realized volatility under hitting times of an irregular grid, Stochastic Processes and their Applications 122 (2012) 3901–3920.
[13] P.R. Hansen, A. Lunde, Realized variance and IID market microstructure noise, Technical report, Dept. of Economics, Brown University, 2004.
[14] T. Hayashi, J. Jacod, N. Yoshida, Irregular sampling and centeral limit theorems for power variations: the continuous case, Annales de l'Institut Henri Poincaré Probabilités et Statistiques 47 (2011) 1197–1218.
[15] J. Jacod, Y. Li, P.A. Mykland, M. Prodolskij, M. Vetter, Microstructure noise in the continuous case: the pre-averaging approach, Stochastic Processes and their Applications 119 (2009) 2249–2276.
[16] J. Jacod, P. Protter, Asymptotic error distributions for the Euler method for stochastic differential equations, The Annals of Probability 26 (1998) 267–307.
[17] I. Kalnina, O. Linton, Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error, Journal of Econometrics 147 (2008) 47–59.
[18] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, Springer, 1991.
[19] Y. Li, P.A. Mykland, Are volatility estimators robust with respect to modeling assumptions? Bernoulli 13 (2007) 601–622.

[20] Y. Li, E. Renault, P.A. Mykland, L. Zhang, X. Zheng, Realized volatility when sampling times are possibly endogenous, Manuscript, 2009. Available at SSRN: http://ssrn.com/abstract=1525410.

[21] Y. Li, Z. Zhang, X. Zheng, Supplement to volatility inference in the presence of both endogenous time and microstructure noise, 2013.

[22] N. Meddahi, E. Renault, B. Werker, GARCH and irregularly spaced data, Economics Letters 90 (2006) 200–204.

[23] P.A. Mykland, L. Zhang, ANOVA for diffusions and Itô processes, Annals of Statistics 34 (2006) 1931–1963.

[24] P.A. Mykland, L. Zhang, Inference for continuous semimartingales observed at high frequency, Econometrica 77 (2009) 1403–1455.

[25] P.A. Mykland, L. Zhang, The econometrics of high frequency data, in: M. Kessler, A. Lindner, M. Sørensen (Eds.), Statistical Methods for Stochastic Differential Equations, Chapman and Hall/CRC Press, 2012, pp. 109–190 (in press).

[26] P.C.B. Phillips, J. Yu, Information loss in volatility measurement with flat price trading, Working Paper, 2007.

[27] M. Podolskij, M. Vetter, Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps, Bernoulli 15 (2009) 634–658.

[28] E. Renault, B.J. Werker, Causality effects in return volatility measures with random times, Journal of Econometrics 160 (2011) 272–279.

[29] D. Revuz, M. Yor, Continuous Martingales and Brownian Motion, Springer-Verlag, Berlin, 1999.

[30] C.Y. Robert, M. Rosenbaum, Volatility and covariation estimation when microstructure noise and trading times are endogenous, Mathematical Finance 22 (2012) 133–164.

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

[32] L. Zhang, From Martingales to ANOVA: implied and realized volatility, Ph.D. Thesis, Department of Statistics, The University of Chicago, 2001.

[33] L. Zhang, Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach, Bernoulli 12 (2006) 1019–1043.

[34] L. Zhang, P.A. Mykland, Y. Aït-Sahalia, A tale of two time scales: determining integrated volatility with noisy high-frequency data, Journal of the American Statistical Association 100 (2005) 1394–1411.