Journal of Econometrics 203 (2018) 187–222

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A unified approach to volatility estimation in the presence of both rounding and random market microstructure noise

Yingying Li [a], Zhiyuan Zhang [b,*], Yichu Li [c]

[a] Hong Kong University of Science and Technology, Hong Kong Special Administrative Region
[b] School of Statistics and Management, Shanghai University of Finance and Economics, No. 777 Guoding Road, Shanghai 200433, China
[c] Investment Technology Group, NY, USA

ARTICLE INFO

Article history:
Received 13 January 2016
Received in revised form 22 November 2017
Accepted 26 November 2017
Available online 6 December 2017

JEL classification:
G12
C22
C14

Keywords:
High-frequency data
Rounding errors
Market microstructure noise
Integrated volatility
Realized volatility

ABSTRACT

Widely used volatility estimation methods mainly consider one of the following two simple microstructure noise models: random additive noise on log prices, or pure rounding errors. Apparently in real data these two types of noise co-exist. In this paper, we discover a common feature of these two types of noise and propose a unified volatility estimation approach in the presence of both rounding and random noise. Our data-driven method enjoys superior properties in terms of bias and convergence rate. We establish feasible central limit theorems and show their superior performance via simulations. Empirical studies show clear advantages of our method when applied to both stocks data and currency exchange data.

1. Introduction

Volatility estimation based on high-frequency data has received great attention over the past decades. A main challenge is the presence of market microstructure noise. Sources of market microstructure noise include various frictions in financial markets such as bid–ask bounce, dealer's inventory control, specific trading mechanism of an exchange and so on. Rounding is also a main source of market microstructure noise because traded prices are all rounded to price grids.

Market microstructure noise accumulates at high frequencies and causes large biases in classical volatility estimators such as realized volatility (RV). This is most clearly seen in the volatility signature plot of Andersen et al. (2000), where RV is plotted against sampling frequencies and an increasing trend of RV as the sampling frequency increases (i.e., the sampling interval decreases) is observed for many stocks.

Great efforts have been made towards consistent and efficient estimators of (integrated) volatilities. Most of the currently widely used high-frequency volatility estimators (see, e.g., the two scales realized volatility (TSRV) by Zhang et al. (2005), the multi-scale realized volatility (MSRV) of Zhang (2006), the realized kernels (RK) by Barndorff-Nielsen et al. (2008), the pre-averaging volatility (PAV) estimator of Jacod et al. (2009), Podolskij and Vetter (2009) and the quasi-maximum likelihood estimator (QMLE) of Xiu, 2010) rely on the assumptions that the market microstructure noise is random and additive on log prices. More precisely, let $(X_t)_{0\le t\le 1}$ be the logarithm of latent price process that follows a semimartingale, then the observed log prices at discrete time points $0 \equiv t_0 < t_1 < \cdots < t_n \equiv 1$ are assumed to be of the form

$$X_{t_i} + \eta_{t_i},\ i = 0, 1, 2, \ldots, n,$$

where $(\eta_{t_i})_{i\ge 0}$ is usually assumed to be independently and identically[1] distributed, and independent of the $(X_t)_{0\le t\le 1}$ process, with $\mathrm{Var}(\eta_{t_i}) = O(1)$. See also Zhou (1996), Hansen and Lunde (2006), Aït-Sahalia et al. (2005) and Bandi and Russell (2006) for related works in this direction.

In Fig. 1, we plot volatility estimates given by the aforementioned random-noise-oriented methods against different sampling intervals for four stocks traded on the New York Stock Exchange

* Corresponding author.
E-mail address: zhang.zhiyuan@mail.shufe.edu.cn (Z. Zhang).

[1] A relaxed condition is assumed in Jacod et al. (2009), see also footnote 2.

https://doi.org/10.1016/j.jeconom.2017.11.006
0304-4076/

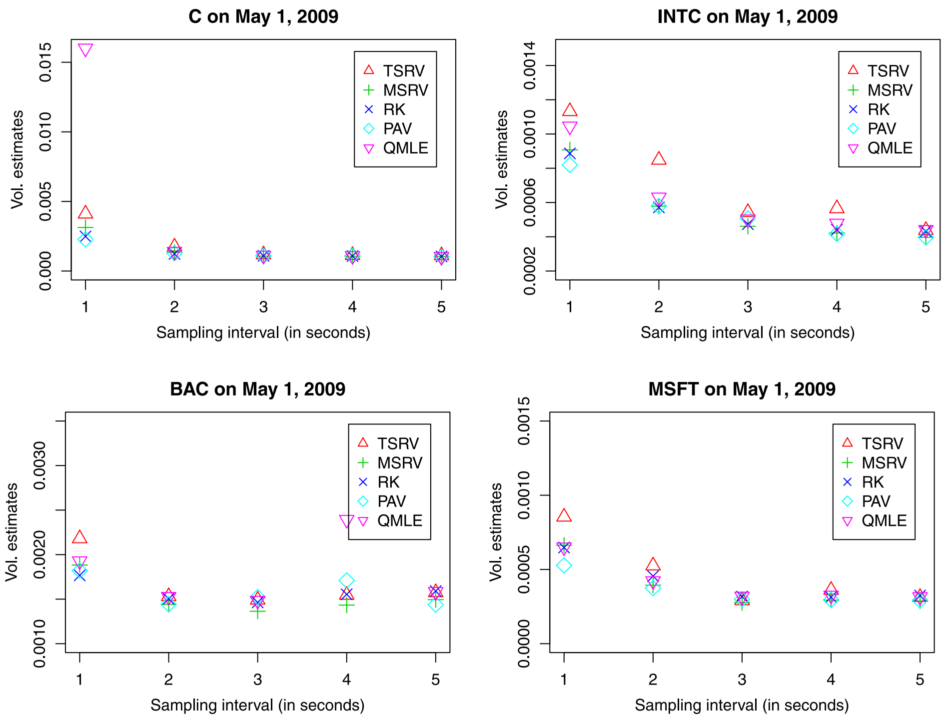

Fig. 1. Volatility estimates of five random-noise-oriented methods across different sampling intervals based on the second-by-second data of four stocks on May 1, 2009. The estimates given by these methods all diverge as the sampling interval becomes smaller, presenting a volatility signature plot pattern.

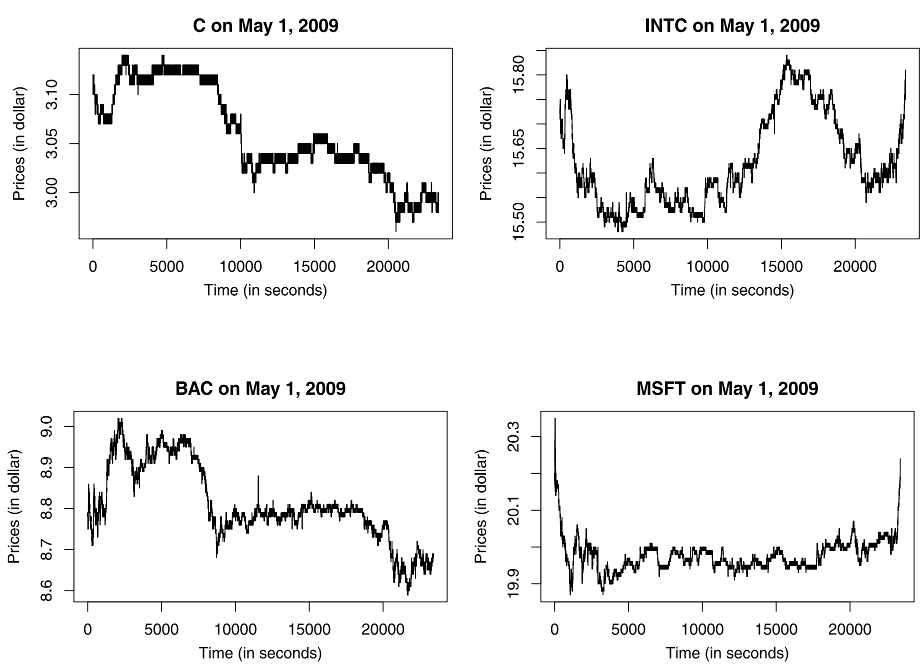

Fig. 2. Second-by-second trade prices of four stocks on May 1, 2009.

(NYSE) and the National Association of Securities Dealers Automated Quotation System (NASDAQ) on May 1, 2009 (Citigroup Inc. (NYSE: C), Intel Corporation (NASDAQ: INTC), Bank of America (NYSE: BAC) and Microsoft Corporation (NASDAQ: MSFT)). These stocks are all traded at high frequencies. Recall as we mentioned above, for many stocks there is an increasing trend of RV when the sampling interval becomes smaller, usually referred to as the "volatility signature plot" pattern. This pattern is undesirable because good volatility estimators should show a converging trend as sampling frequency goes higher rather than diverging. One would expect the random-noise-oriented methods mentioned above would no longer exhibit this signature plot pattern. However, in Fig. 1, we see that the widely used volatility estimators, TSRV, MSRV, RK, PAV and QMLE, still show volatility signature plot pattern.

What has gone wrong with these volatility estimators? The answer has to do with the fact that the observed high-frequency prices are also exposed to rounding errors. Li and Mykland (2007) show that ignoring rounding errors may lead to large biases in volatility estimation. We plot in Fig. 2 the second-by-second price processes of the four stocks studied in Fig. 1. The prices of all these stocks are rounded to one cent. At first glance, we see that rounding is most apparent in the prices of Citigroup Inc. The presence of rounding can also be seen for the other three stocks although much less apparent.

Rounding is an inevitable source that affects the performance of the widely used volatility estimators. However, in the literature, less attention has been paid to the rounding effect. Studies on volatility estimation with rounded prices can be traced back to Delattre and Jacod (1997). More recent works include the seminal papers Rosenbaum (2009) and Li and Mykland (2015) among others. However, these papers only focus on the case when there is pure rounding which is also unrealistic.

Is it possible to find a unified volatility estimation method that works well in the presence of both rounding errors and random market microstructure noise?

In this paper, we propose such a method that allows the market microstructure noise to be a mixture of random additive noise and rounding. To elaborate our idea, we make a "small-error assumption" which helps bridge the two types of noise that are of distinct nature. In particular, we find a common feature shared by the biases caused by these two types of noise, based on which we build our estimation method. We establish feasible central limit theorems which provide the theoretical basis for interval inference and beyond. Our proposed approach not only enjoys consistency with both types of noise incorporated but also reaches towards the best possible convergence rate. In addition, we provide a data-adaptive automated procedure for selecting parameter values involved in the proposed estimator, thus, each individual price process uses a set of parameter values that are tailor-made to itself towards an unbiased efficient volatility estimate.

To be more specific, we assume that the logarithm of latent price process follows the following Itô diffusion process

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

where $(\mu_t)_{0\le t\le 1}$ and $(\sigma_t)_{0\le t\le 1}$ are locally bounded and adapted random processes and $(W_t)_{0\le t\le 1}$ is a standard Brownian motion. Hence, the latent price process $(S_t)_{0\le t\le 1}$ ($S_t := \exp(X_t)$) satisfies

$$dS_t = \left(\mu_t + \frac{\sigma_t^2}{2}\right) S_t dt + \sigma_t S_t dW_t. \tag{2}$$

The contaminated price process $(\breve{S}_t)_{0\le t\le 1}$ is observed at discrete time points $0 \equiv t_0 < t_1 < \cdots < t_n \equiv 1$:

$$\breve{S}_{t_i} = \left(e^{X_{t_i}+\eta_{t_i}}\right)^{(\alpha_n)},\ i = 0, 1, 2, \ldots, n, \tag{3}$$

where for any $s > 0$, $s^{(\alpha_n)} := \alpha_n \vee ([s/\alpha_n]\alpha_n)$ (i.e., rounded to the nearest multiples of tick size α_n), and $(\eta_{t_i})_{i\ge 0}$ is a sequence of random additive noises with common standard deviation γ_n. Models of a similar form as (3) have also been considered in Li and Mykland (2007), Jacod et al. (2009) and Jacod et al. (2017a). The "small-error assumption" that will be relied on in developing our asymptotic theory refers to that the sizes of random noise and rounding, i.e., γ_n and α_n, both shrink as the observation frequency n increases. Although we only observe one n, one α_n and one γ_n, we can always develop an asymptotic theory based on the assumption that both α_n and γ_n shrink with n and then apply the theory on the actual noise sizes and observation frequency. On the empirical side, the small-error assumption does appear to be sensible. For example, it is believed in Zhang et al. (2005) that γ_n being around 0.0005 is a reasonable value for the size of the random additive noise. This means that $\gamma_n^2 * n$ (around 0.00585 when $n = 23{,}400$) is still quite small or, in other words, we can say that γ_n^2 is comparable to $1/n^{\iota}$ for some $\iota \ge 1$. See also Hansen and Lunde (2006, p. 154) for more empirical evidence that the market microstructure noise is small. Related literature that considers small-error includes Kolassa and McCullagh (1990), Delattre and Jacod (1997), Gloter and Jacod (2001a, b), Barndorff-Nielsen et al. (2008, p. 1501), Rosenbaum (2009), Li et al. (2016) and Li and Mykland (2015) among others.

Our unified approach is based on an interesting observation: when both rounding and random additive noise are present, in the cases when either rounding dominates (see Condition RO(a)-(b) in Section 2) or random noise dominates (see Condition RA(a)-(b) in Section 2), the leading bias terms in RV share a same characteristic, which is that they are both in the form of RV sampling frequency multiplied by a factor that only depends on the latent price process and the noise level. Such a feature allows us to find a unified volatility (UV) estimator, see Section 3 for details.

We close this section by summarizing our contributions as follows.

- Our unified volatility estimator is the first[2] method that can achieve consistent estimates of integrated volatility when both rounding and general random microstructure noise are present.
- We discover a common feature of the biases in RV due to rounding and random additive noise, and exploit the common feature to build our estimator. This is the first paper to document such an interesting link, and demonstrate an idea to overcome the challenges coming from noises of distinct nature. Our method turns out to have a simple regression format.[3]
- We propose a data-driven method to decide what parameter values to use in our estimator. This procedure helps identify the right asymptotic scheme which leads to high convergence rate. As an example, when the automated data-driven procedure identifies that the data can be viewed as satisfying Condition RA(b') or Condition RO(b') as specified in Section 3.4, a $\sqrt{n}$ rate of convergence can be achieved.
- Our feasible central limit theorems are tested via extensive simulation studies. Compared with well established random-noise-oriented methods and rounding-error-oriented methods (see Section 4), our method consistently produces better point and interval estimates, for cases ranging from when rounding effect can be negligible to when rounding becomes the main component of the noise. Our theory was also tested to a case where the theoretical assumptions are not fully satisfied, and our estimator is seen to be robust.
- Empirical studies exhibit clear advantage of our method. Our method yields stable estimates across different sampling frequencies, whereas methods under comparison often exhibit apparent biases at high frequencies. This is seen on both stocks data and currency exchange data.

The rest of the paper is organized as follows. Section 2 gives basic assumptions and some important preliminary results. With the results developed in Section 2 as the basis, we introduce the UV estimator, the intuition behind it and its asymptotic properties in Section 3, where a practical data-adaptive procedure is proposed.

---

2 Notice that Jacod et al. (2009) covers a special random noise model with rounding. In fact, as remarked in Li and Mykland (2007), what is essentially needed is $E(Y_{t_i}|X_{t_i}) = X_{t_i}$, where Y_{t_i} stands for the observed log price at time t_i (see equation (2.3) on page 2252 of Jacod et al., 2009). In certain special cases involving rounding (such as model 2 on page 2257 of Jacod et al., 2009), this condition is met. In those special situations the noise can be dealt with in the same way as simple additive noise. In this paper, we do not require such a condition.

3 This regression format may look similar to MSRV by Zhang (2006). We emphasize that due to the additional consideration of rounding, our method uses different regression schemes than MSRV and the asymptotics are also significantly different. In particular, when used on data involve rounding, MSRV is asymptotically biased (see reasons in Li and Mykland, 2007) while UV is not. See also simulation and empirical studies for numerical performance.

A comprehensive simulation study is presented in Section 4 where UV is compared with the widely used estimators. Empirical study is conducted in Section 5 showing clear advantages of our method in practice. Proofs are collected in Appendix.

2. Preliminaries

In this section, we present preliminary results that lead to the idea of a unified approach.

2.1. Setup

We state an overall assumption first.

Assumption 1. The latent log price and the observed price follow (1) and (3). There is a filtration $(\mathcal{F}_t)_{0\le t\le 1}$ with respect to which $(W_t)_{0\le t\le 1}$, $(\mu_t)_{0\le t\le 1}$ and $(\sigma_t)_{0\le t\le 1}$ in (1) are adapted. The volatility process $(\sigma_t)_{0\le t\le 1}$ is continuous and satisfies $\inf_{t\in[0,1]}\sigma_t > 0$. Observation times follow $t_i = i/n$, $i = 0, 1, 2, \ldots, n$. The random noise sequence $(\eta_{t_i})_{i\ge 0}$ consists of i.i.d. Normal$(0, \gamma_n^2)$ random variables[4] that are independent of $\mathcal{F}_1$. Furthermore, the filtration $(\mathcal{F}_t)_{0\le t\le 1}$ is generated by finitely many continuous martingales.

The following conditions concern the relation between α_n and γ_n. For clarity, we state them equivalently based on the relation between $\widetilde{\alpha}_n$ and γ_n, where

$$\widetilde{\alpha}_n = \frac{\alpha_n}{S_0}$$

represents the size of rounding errors on log scale,[5] which can be directly compared with γ_n, the size of the random additive noise on log prices.

We shall establish theories when either one of the following two conditions is satisfied. That is, either random microstructure noise dominates or rounding effect dominates. Let $0 < \kappa < 1$ be such that either Condition RA(a)-(b) or Condition RO(a)-(b) holds:

Condition RA (*Random noise dominates*).

(a) $\widetilde{\alpha}_n/\gamma_n \to 0$ *and* $(n^{1-\kappa})^{1/2}\gamma_n \to \gamma_\kappa \in [0,\infty)$*; or a bit more strongly,*
(b) $(n^{1-\kappa})^{1/2}\widetilde{\alpha}_n/\gamma_n \to 0$ *and* $n^{1-\kappa}\gamma_n^2 = \gamma_\kappa^2 + o(1)$ *for some* $\gamma_\kappa \in [0,\infty)$.

Condition RO (*Rounding effect dominates*).

(a) $\gamma_n/\widetilde{\alpha}_n \to 0$ *and* $(n^{1-\kappa})^{1/2}\widetilde{\alpha}_n \to \beta_\kappa \in [0,\infty)$*; or a bit more strongly,*
(b) $(n^{1-\kappa})^{1/2}\gamma_n/\widetilde{\alpha}_n \to 0$ *as a power law function in n and* $n^{1-\kappa}\widetilde{\alpha}_n{}^2 = O\left(1/(n^{1-\kappa})^{1/2}\right)$*;*
(c) $\sigma_t = \sigma(S_t)$ *is a function of* S_t *and the function* $\sigma(\cdot)$ *is of class* C^5 *on* $[0,\infty)$*;*
(d) *The diffusion* $(S_t)_{0\le t\le 1}$ *in* (2) *admits transition density* $p_{S_t|S_u}(y|x)$ *for* $0 \le u < t \le 1$*. Moreover, there exists a universal constant* $C > 0$ *such that, for* $0 \le u < t \le 1$*,*

$$\sup_{y,x\in[0,\infty)} p_{S_t|S_u}(y|x) \le \frac{C}{\sqrt{t-u}}.$$

[4] The normality assumption here is adopted from Li and Mykland (2007) as it facilitates reasonably explicit computations. Mild relaxation of this assumption on the density and moment conditions may lead to similar results but is beyond the scope of this paper.

[5] Notice that the rounding error at $t = 0$ on log scale is $\log(S_0^{(\alpha_n)}) - \log(S_0) = \log(1+(S_0^{(\alpha_n)}-S_0)/S_0) \approx (S_0^{(\alpha_n)}-S_0)/S_0$, whose size can be described by $\widetilde{\alpha}_n = \alpha_n/S_0$.

Remark 1. Condition RO(c), which is required for roughly half of the results in this paper, is adopted from Delattre and Jacod (1997) and Li and Mykland (2015). A formal extension of this condition to the case of σ_t being general stochastic processes is beyond the scope of this paper. However, we show in simulation studies that our proposed estimator works well for cases where Condition RO(c) is violated (i.e., when σ_t is not a function of S_t).

Remark 2. Our asymptotic framework, i.e., Condition RA(a)-(b) (resp. Condition RO(a)-(b)), allows for the case $n\gamma_n^2 \to \infty$ (resp. $n\widetilde{\alpha}_n{}^2 \to \infty$). That is, the raw RV is allowed to have a divergent bias.

We denote the complete grid of observation times as $\mathcal{G} := \{t_0, t_1, \ldots, t_n\}$ and the logarithm of the observed prices as

$$Y_{t_i} := \log\left(\breve{S}_{t_i}\right),\ i = 0, 1, 2, \ldots, n,$$

where $\breve{S}_{t_i}$ is as defined in (3). The RV of $(Y_t)_{0\le t\le 1}$ over $\mathcal{G}$ is defined as

$$[Y,Y]_1^{(\mathrm{all})} := \sum_{i=0}^{n-1}(\Delta Y_{t_i})^2,$$

where $\Delta Y_{t_i} = Y_{t_{i+1}} - Y_{t_i}$. The target of interest is the integrated volatility over the time horizon $[0, 1]$:

$$\langle X, X\rangle_1 := \int_0^1 \sigma_t^2 dt.$$

Let

$$\mathcal{G}_s := \{\tau_0, \tau_1, \ldots, \tau_{n_s}\}$$

be a generic sub-grid of equally spaced times which are taken every $[n^\kappa]$th point from the complete grid $\mathcal{G}$ of times starting from one of the first $[n^\kappa]$ time points, where the exponent κ is the same as in Condition RA(a)-(b) and Condition RO(a)-(b). Then the number of observations n_s over sub-grid $\mathcal{G}_s$ satisfies

$$n_s = n^{1-\kappa} + O(1).$$

Similarly as in the complete grid case, we define the RV of $(Y_t)_{0\le t\le 1}$ over $\mathcal{G}_s$ as

$$[Y,Y]_1^{\mathcal{G}_s} := \sum_{i=0}^{n_s-1}(\Delta Y_{\tau_i})^2,$$

where $\Delta Y_{\tau_i} = Y_{\tau_{i+1}} - Y_{\tau_i}$. Next, we shall provide preliminary convergence results over such sub-grid of times $\mathcal{G}_s$.

2.2. When rounding effect dominates

Proposition 1. *If Assumption 1, Condition RO(a) and RO(c)-(d) are satisfied, then*

$$[Y,Y]_1^{\mathcal{G}_s} - \int_0^1 \sigma_t^2 dt - \frac{n_s\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt + \underbrace{\frac{n_s\alpha_n^2}{\pi^2}\int_0^1 \frac{1}{S_t^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\exp\left(-2\pi^2k^2\frac{\sigma_t^2S_t^2}{n_s\alpha_n^2}\right)dt}_{\mathrm{NLB}(n_s)} \xrightarrow{p} 0. \tag{4}$$

We use notation NLB(n_s) in (4) to denote the bias due to rounding that is nonlinear in n_s.

Proposition 2. *If Assumption 1 and Condition RO(b)-(d) are satisfied, then stably in law,*

$$\sqrt{n^{1-\kappa}}\left([Y,Y]_1^{\mathcal{G}_s} - \int_0^1 \sigma_t^2 dt - \frac{n_s\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt\right) \to \int_0^1 \sqrt{2}\sigma_t^2 dB_t,$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion that is independent of the underlying σ-field.

2.3. When random noise dominates

Proposition 3. *If Assumption 1 and Condition RA(a) are satisfied, then*

$$[Y,Y]_1^{\mathcal{G}_s} - \int_0^1 \sigma_t^2 dt - 2n_s\gamma_n^2 \xrightarrow{p} 0.$$

Proposition 4. *If Assumption 1 and Condition RA(b) are satisfied, then stably in law,*

$$\sqrt{n^{1-\kappa}}\left([Y,Y]_1^{\mathcal{G}_s} - \int_0^1 \sigma_t^2 dt - 2n_s\gamma_n^2\right) \to \int_0^1 \left(2\sigma_t^4 + 8\gamma_\kappa^2\sigma_t^2 + 12\gamma_\kappa^4\right)^{\frac{1}{2}} dB_t,$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion that is independent of the underlying σ-field.

2.4. A unified proposition

The following Proposition 5 unifies the results in Sections 2.2 and 2.3.

Proposition 5. *If Assumption 1 and either Condition RA(b) or Condition RO(b)-(d) are satisfied, then stably in law,*

$$\sqrt{n^{1-\kappa}}\left([Y,Y]_1^{\mathcal{G}_s} - \int_0^1 \sigma_t^2 dt - \underbrace{2n_s\gamma_n^2}_{\mathrm{LB_{ra}}(n_s)} - \underbrace{\frac{n_s\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt}_{\mathrm{LB_{ro}}(n_s)}\right) \to \int_0^1 \left(2\sigma_t^4 + 8\gamma_\kappa^2\sigma_t^2 + 12\gamma_\kappa^4\right)^{\frac{1}{2}} dB_t, \quad (5)$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion that is independent of the underlying σ-field and $\gamma_\kappa = 0$ under Condition RO(b).

We refer to the first part of the bias $2n_s\gamma_n^2$ due to random noise as "$\mathrm{LB_{ra}}(n_s)$"; and refer to the second part of the bias $n_s\alpha_n^2/6\int_0^1 1/S_t^2 dt$ due to rounding as "$\mathrm{LB_{ro}}(n_s)$".

Remark 3. When comparing Proposition 2 with Proposition 4, we can see that the leading bias terms in the two cases share a same characteristic, which is that they are both in the form of n_s multiplied by a factor that only depends on the latent price process and the noise level. Proposition 5 unifies the two cases and clearly shows that the bias $\mathrm{LB_{ra}}(n_s) + \mathrm{LB_{ro}}(n_s)$ is linear in n_s. This fact facilitates the development of the unified approach presented in Section 3.

Remark 4. In the case $\kappa = 0$, which we shall discuss in details in Section 3.4, it follows easily that the previous preliminary convergence results also hold with $[Y,Y]_1^{\mathcal{G}_s}$ and n_s being replaced by $[Y,Y]_1^{(\mathrm{all})}$ and n, respectively. Otherwise, we select the right asymptotic scheme by using such appropriate sub-grid as $\mathcal{G}_s$ where κ satisfies Condition RO(a) (resp. RO(b)) so that the bias due to rounding just goes down to the level that makes the estimation (resp. inference) of volatility possible. Although TSRV, MSRV, PAV, RK etc. also involve subsampling, ignoring rounding leads to different subsampling schemes that yield biased estimates when rounding does exist.

3. The UV estimator

In this section, we show how the feature found in Proposition 5 can be exploited towards a unified volatility estimation procedure. Notice that we only make use of one single sub-grid and throw away a lot of observations in developing the preliminary results in Section 2. In the following, we shall make use of all observations and achieve higher efficiency.

First, we introduce some additional notation. For any fixed L, we choose L *sequences* of positive integers $K_1 < K_2 < \cdots < K_L$ satisfying

$$K_l = \mathcal{C}_l n^\kappa + O(1), \quad l = 1, 2, \ldots, L, \quad (6)$$

where $0 < \mathcal{C}_1 < \cdots < \mathcal{C}_L$ are constants and κ is the same as in Condition RA(a)-(b) and Condition RO(a)-(b). Condition RA(a)-(b) or Condition RO(a)-(b) can be met by choosing an appropriate $\kappa > 0$ or equivalently, the subsample step sizes K_l's according to (6). Naturally, in this procedure, κ (or equivalently, K_l's) should be chosen to be as small as possible to achieve a high convergence rate. We will discuss later that in practice we can select K_l's in a data-driven way.

Define sub-grids

$$\mathcal{G}_{k,l} := \{t_{k+iK_l} \in \mathcal{G} : i = 0, 1, 2, \ldots\},$$
$$\text{where } k = 0, 1, \ldots, K_l - 1 \text{ and } l = 1, 2, \ldots, L.$$

That is, $\mathcal{G}_{k,l}$ is a collection of observation times by taking every K_lth point from the complete grid $\mathcal{G}$ starting from t_k. Let

$$[Y,Y]_1^{\mathcal{G}_{k,l}} := \sum_{t_i\in\mathcal{G}_{k,l}} (Y_{t_{i,+}} - Y_{t_i})^2$$

denote the RV of $(Y_t)_{0\leq t\leq 1}$ over the sub-grid $\mathcal{G}_{k,l}$, where if $t_i \in \mathcal{G}_{k,l}$, then $t_{i,+}$ denotes the next time in $\mathcal{G}_{k,l}$. Let

$$\bar{n}_l := \frac{n}{K_l}$$

be the average grid size, and let $[Y,Y]_1^{(\mathrm{avg},l)}$ be the averaged RV over K_l sub-grids $\mathcal{G}_{k,l}$, $k = 0, 1, \ldots, K_l - 1$, namely,

$$[Y,Y]_1^{(\mathrm{avg},l)} := \frac{1}{K_l}\sum_{k=0}^{K_l-1} [Y,Y]_1^{\mathcal{G}_{k,l}}. \quad (7)$$

As an extension of Proposition 5, we have the following Proposition 6.

Proposition 6. *Under Assumption 1, for any fixed L, if the L sequences of positive integers $K_1 < K_2 < \cdots < K_L$ are chosen according to (6) and $\kappa > 0$ is chosen to be one such that either Condition RA(b) or Condition RO(b)-(d) hold, then stably in law,*

$$\sqrt{n^{1-\kappa}}\begin{pmatrix} [Y,Y]_1^{(\mathrm{avg},1)} - \int_0^1 \sigma_t^2 dt - 2\bar{n}_1\gamma_n^2 - \frac{\bar{n}_1\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt \\ [Y,Y]_1^{(\mathrm{avg},2)} - \int_0^1 \sigma_t^2 dt - 2\bar{n}_2\gamma_n^2 - \frac{\bar{n}_2\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt \\ \vdots \\ [Y,Y]_1^{(\mathrm{avg},L)} - \int_0^1 \sigma_t^2 dt - 2\bar{n}_L\gamma_n^2 - \frac{\bar{n}_L\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt \end{pmatrix} \to \begin{pmatrix} \widetilde{N}_1 \\ \widetilde{N}_2 \\ \vdots \\ \widetilde{N}_L \end{pmatrix},$$

where $(\widetilde{N}_1, \ldots, \widetilde{N}_L)' \stackrel{d}{=} \widetilde{\mathbf{Q}}^{1/2}\widetilde{\mathbf{Z}}$, $\widetilde{\mathbf{Z}}$ is a standard L-dimensional Normal random vector that is independent of the underlying σ-field, and

$$\widetilde{\mathbf{Q}} = (\tilde{q}_{l_1,l_2})_{1\leq l_1,l_2\leq L} \quad (8)$$

with $\tilde{q}_{l_2,l_1} = \tilde{q}_{l_1,l_2}$ *and*

$$\tilde{q}_{l_1,l_2} = \frac{2\mathcal{C}_{l_1}(3\mathcal{C}_{l_2} - \mathcal{C}_{l_1})}{3\mathcal{C}_{l_2}} \int_0^1 \sigma_t^4 dt \quad \text{for } 1 \le l_1 \le l_2 \le L. \tag{9}$$

3.1. Towards a unified approach

Proposition 6 conveys the following message,

$$\left.\begin{aligned} &[Y,Y]_1^{(\text{avg},1)} - \int_0^1 \sigma_t^2 dt - \bar{n}_1\left(2\gamma_n^2 + \frac{\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt\right) \\ &\qquad = \frac{1}{\sqrt{n^{1-\kappa}}}\widetilde{N}_1 + o_p\left(\frac{1}{\sqrt{n^{1-\kappa}}}\right) \\ &[Y,Y]_1^{(\text{avg},2)} - \int_0^1 \sigma_t^2 dt - \bar{n}_2\left(2\gamma_n^2 + \frac{\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt\right) \\ &\qquad = \frac{1}{\sqrt{n^{1-\kappa}}}\widetilde{N}_2 + o_p\left(\frac{1}{\sqrt{n^{1-\kappa}}}\right) \\ &\qquad \vdots \\ &[Y,Y]_1^{(\text{avg},L)} - \int_0^1 \sigma_t^2 dt - \bar{n}_L\left(2\gamma_n^2 + \frac{\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt\right) \\ &\qquad = \frac{1}{\sqrt{n^{1-\kappa}}}\widetilde{N}_L + o_p\left(\frac{1}{\sqrt{n^{1-\kappa}}}\right) \end{aligned}\right\}, \tag{10}$$

where $(\widetilde{N}_1, \widetilde{N}_2, \ldots, \widetilde{N}_L)'$ is multivariate Normal (conditional on $\mathcal{F}_1$) with mean zero and covariance matrix $\widetilde{\mathbf{Q}}$ given in (8) and (9).

Inspired by (10), we propose to consider the following linear model, which depicts the (approximate) linear relation between the averaged RV, $[Y,Y]_1^{(\text{avg},l)}$, and the average grid size $\bar{n}_l$,

$$\mathbb{Y} = \mathbb{X}\mathbf{b} + \mathbf{e}, \tag{11}$$

where

$$\mathbb{Y} = \begin{pmatrix} [Y,Y]_1^{(\text{avg},1)} \\ \vdots \\ [Y,Y]_1^{(\text{avg},L)} \end{pmatrix}, \quad \mathbb{X} = \begin{pmatrix} 1 & \bar{n}_1 \\ 1 & \bar{n}_2 \\ \vdots & \vdots \\ 1 & \bar{n}_L \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \int_0^1 \sigma_t^2 dt \\ \frac{\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt + 2\gamma_n^2 \end{pmatrix}$$

and the error term $\mathbf{e}$ is an $L \times 1$ random vector that is asymptotically multivariate Normal (conditional on $\mathcal{F}_1$) with mean zero and covariance matrix $\widetilde{\mathbf{Q}}/n^{1-\kappa}$.

Since our target $\int_0^1 \sigma_t^2 dt$ is part of the coefficient in the regression equation (11), the ordinary least squares (OLS) estimation method can be applied. More specifically, an estimator for the coefficient vector $\mathbf{b} := (b_0, b_1)'$ is

$$\widehat{\mathbf{b}} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}.$$

In particular, an estimator for b_0, i.e., $\int_0^1 \sigma_t^2 dt$, is

$$\widehat{\langle X, X\rangle}_1 := \sum_{l=1}^{L} \omega_l [Y,Y]_1^{(\text{avg},l)}, \tag{12}$$

where

$$\omega_l := \frac{1}{L} - \bar{\bar{n}}\frac{\bar{n}_l - \bar{\bar{n}}}{\sum_{j=1}^{L}\bar{n}_j^2 - L\bar{\bar{n}}^2} \text{ and } \bar{\bar{n}} := \frac{1}{L}\sum_{l=1}^{L}\bar{n}_l. \tag{13}$$

We call the above estimator (12) the *unified volatility* (UV as previously abbreviated) estimator.

Remark 5. About the positivity of UV, our data-driven procedure to select the parameters is very helpful in this respect but is still a bit ad hoc. We will not pursue formal establishment of almost sure positivity of the estimator. We emphasize, however, in our extensive numerical studies, UV had never produced negative estimates.

Remark 6. Instead of using OLS, one can use generalized least squares (GLS) method to, possibly, achieve higher efficiency. However, better efficiency of GLS is only guaranteed when one knows the explicit covariance matrix of the error term $\mathbf{e}$, which is not the case here. The finite sample property of the feasible GLS estimator varies dramatically with each particular model (see Chapter 3 of Strutz, 2010). In our unreported simulation studies, the feasible GLS estimator gives larger finite sample variance than the OLS estimator. Hence, we do not pursue further along this path.

3.2. Central limit theorem for the UV estimator

In this section, we provide the asymptotic property of $\widehat{\langle X, X\rangle}_1$ given by (12). For ω_l defined in (13), it can be easily seen that $\sum_{l=1}^{L}\omega_l = 1$, $\sum_{l=1}^{L}\omega_l\bar{n}_l = 0$ and $\omega_l = O(1)$. These facts together with (10) lead to the following decomposition for UV estimator

$$\widehat{\langle X, X\rangle}_1 = \int_0^1 \sigma_t^2 dt + \frac{1}{\sqrt{n^{1-\kappa}}}\sum_{l=1}^{L}\omega_l\widetilde{N}_l + o_p\left(\frac{1}{\sqrt{n^{1-\kappa}}}\right). \tag{14}$$

Based on this, the variance of the estimator $\widehat{\langle X, X\rangle}_1$ is approximated by

$$\frac{1}{n^{1-\kappa}}\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\omega_{l_1}\omega_{l_2}\tilde{q}_{l_1,l_2}, \tag{15}$$

where $\omega_{l_i}, i = 1, 2$, are given by (13) and $\tilde{q}_{l_1,l_2}$ is given by (8) and (9). More specifically, we have the following central limit theorem.

Theorem 1. *Under the conditions of Proposition 6, stably in law,*

$$\sqrt{\frac{n}{K_1}}\left(\widehat{\langle X, X\rangle}_1 - \int_0^1 \sigma_t^2 dt\right) \to \sqrt{\widetilde{\mathcal{V}}}\widetilde{N},$$

where $\widetilde{N}$ is a standard Normal random variable that is independent of the underlying σ-field and

$$\widetilde{\mathcal{V}} =: (1/\mathcal{C}_1)\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\tilde{q}_{l_1,l_2}\breve{\omega}_{l_1}\breve{\omega}_{l_2},$$

$$\text{where } \breve{\omega}_l := \frac{1}{L} - \frac{\frac{1}{L}\sum_{j=1}^{L}\frac{1}{\mathcal{C}_j}\left(\frac{1}{\mathcal{C}_l} - \frac{1}{L}\sum_{j=1}^{L}\frac{1}{\mathcal{C}_j}\right)}{\sum_{j=1}^{L}\frac{1}{\mathcal{C}_j^2} - L\left(\frac{1}{L}\sum_{j=1}^{L}\frac{1}{\mathcal{C}_j}\right)^2}. \tag{16}$$

3.3. A feasible central limit theorem

To make Theorem 1 feasible, we need to estimate the quantity $\int_0^1 \sigma_t^4 dt$. For this, we need the following notation. Define $[Y,Y,Y,Y]_1^{(\text{all})} := \sum_{i=0}^{n-1}(\Delta Y_{t_i})^4$, where $\Delta Y_{t_i} = Y_{t_{i+1}} - Y_{t_i}$. Similarly, given any positive integer K that may diverge with n, for $k = 0, 1, \ldots, K-1$, define $[Y,Y,Y,Y]_1^{\mathcal{G}_K^{(k)}} := \sum_{t_i \in \mathcal{G}_K^{(k)}}(Y_{t_{i,+}} - Y_{t_i})^4$, where $\mathcal{G}_K^{(k)} := \{t_{k+iK} \in \mathcal{G} : i = 0, 1, 2, \ldots\}$ and $t_{i,+}$ is the time that follows t_i in $\mathcal{G}_K^{(k)}$. We further define

$$\mathcal{Q} := \frac{2n}{3K^2}\sum_{k=0}^{K-1}[Y,Y,Y,Y]_1^{\mathcal{G}_K^{(k)}}. \tag{17}$$

Proposition 7. *Under Assumption 1.*

(1). If either (a) $\widetilde{\alpha_n}/\gamma_n \to 0$ and $\sqrt{n}\gamma_n \to 0$ or (b) $\gamma_n/\widetilde{\alpha_n} \to 0$, $\sqrt{n}\widetilde{\alpha_n} \to 0$ and Condition RO(d) hold, then

$$\frac{2n}{3}[Y,Y,Y,Y]_1^{(\text{all})} \xrightarrow{p} 2\int_0^1 \sigma_t^4 dt.$$

(2). *Suppose that either (a) $\widetilde{\alpha}_n/\gamma_n \to 0$ and $\sqrt{n/K}\gamma_n \to 0$ or (b) $\gamma_n/\widetilde{\alpha}_n \to 0$, $\sqrt{n/K}\widetilde{\alpha}_n \to 0$ and Condition RO(d) hold, then*

$$\mathcal{Q} \xrightarrow{p} 2\int_0^1 \sigma_t^4 dt.$$

We can then define an estimate of the asymptotic variance $\widetilde{\mathcal{V}}$ as follows

$$\widehat{\widetilde{\mathcal{V}}} := \left(\frac{1}{\mathcal{C}_1}\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\omega_{l_1}\omega_{l_2}h_{l_1,l_2}\right)\frac{\mathcal{Q}}{2}, \tag{18}$$

where ω_{l_i}, $i = 1, 2$, are given by (13), and h_{l_1,l_2}'s with $h_{l_2,l_1} = h_{l_1,l_2}$ are given as follows

$$h_{l_1,l_2} = \frac{2\mathcal{C}_{l_1}(3\mathcal{C}_{l_2} - \mathcal{C}_{l_1})}{3\mathcal{C}_{l_2}} \quad \text{for } 1 \le l_1 \le l_2 \le L.$$

We have the following feasible central limit theorem.

Corollary 1. *Under the conditions of Theorem 1, if $K_L/K = o(1)$ under Condition RA(b) or $K_L/K = O(1)$ under Condition RO(b), then stably in law,*

$$\sqrt{\frac{n}{K_1\widehat{\widetilde{\mathcal{V}}}}}\left(\widehat{\langle X, X\rangle}_1 - \int_0^1 \sigma_t^2 dt\right) \to \widetilde{Z},$$

where $\widetilde{Z}$ is a standard Normal random variable that is independent of the underlying σ-field, and $\widehat{\widetilde{\mathcal{V}}}$ is as defined in (18).

The beginning of Sections 3.5 and 3.5.1 provide detailed information about choosing K_l's. In Section 3.5.2, we discuss how to choose K in practice.

3.4. The case when $\kappa = 0$

In this section, we provide analogous results in the case $\kappa = 0$, i.e., when data allows one to achieve a possible $\sqrt{n}$ rate of convergence. To formalize our procedure, we state the counterparts of Condition RA(b) and Condition RO(b) in the case $\kappa = 0$ as follows.

Condition RA (*b'*). *$\sqrt{n}\widetilde{\alpha}_n/\gamma_n \to 0$ and $n\gamma_n^2 = \gamma^2 + o(1)$ for some $\gamma \in [0, \infty)$.*

Condition RO (*b'*). *$\sqrt{n}\gamma_n/\widetilde{\alpha}_n \to 0$ as a power law function in n and $n\widetilde{\alpha}_n^2 = O\left(1/n^{1/2}\right)$.*

Then for any fixed L, we choose L *fixed* positive integers $K_1 < K_2 < \cdots < K_L$. Let again $\bar{n}_l := n/K_l$, $l = 1, 2, \ldots, L$, be average grid sizes and define $[Y, Y]_1^{(\text{avg},l)}$, $l = 1, 2, \ldots, L$, the same way as in (7) with K_l's being now fixed integers. To echo the discussion at the beginning of Section 3, in meeting Condition RO(b) or Condition RA(b), we should choose as small as possible a value for κ (or equivalently, K_l's according to (6)) to achieve a high convergence rate. The highest possible rate that we can achieve is as induced by Condition RA(b') or Condition RO(b') depending on the nature of data. We shall in the next subsection discuss how one can in practice select K_l's in a data-driven way.

When K_l's are fixed rather than being chosen as in (6) with $\kappa > 0$, analogous results to that in Proposition 6 is given as follows.

Proposition 8. *Under Assumption 1, for any fixed L and fixed positive integers $K_1 < K_2 < \cdots < K_L$, if either Condition RA(b') or Condition RO(b') and RO(c)-(d) hold, then stably in law,*

$$\sqrt{n}\begin{pmatrix} [Y,Y]_1^{(\text{avg},1)} - \int_0^1 \sigma_t^2 dt - 2\bar{n}_1\gamma_n^2 - \frac{\bar{n}_1\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt \\ [Y,Y]_1^{(\text{avg},2)} - \int_0^1 \sigma_t^2 dt - 2\bar{n}_2\gamma_n^2 - \frac{\bar{n}_2\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt \\ \vdots \\ [Y,Y]_1^{(\text{avg},L)} - \int_0^1 \sigma_t^2 dt - 2\bar{n}_L\gamma_n^2 - \frac{\bar{n}_L\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt \end{pmatrix} \to \begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ N_L \end{pmatrix},$$

where $(N_1, \ldots, N_L)' \stackrel{d}{=} \mathbf{Q}^{1/2}\mathbf{Z}$, $\mathbf{Z}$ is a standard L-dimensional Normal random vector that is independent of the underlying σ-field, and

$$\mathbf{Q} = (q_{l_1,l_2})_{1\le l_1,l_2\le L} \tag{19}$$

with $q_{l_2,l_1} = q_{l_1,l_2}$ and

$$\left.\begin{aligned} q_{l,l} &= \frac{12\gamma^4}{K_l^2} + \frac{8\gamma^2}{K_l}\int_0^1 \sigma_t^2 dt \\ &\quad + \left(1 + \frac{(K_l-1)(2K_l-1)}{3K_l}\right)\int_0^1 2\sigma_t^4 dt \quad \text{for } 1 \le l \le L, \\ q_{l_1,l_2} &= \frac{8\gamma^4}{K_{l_1}K_{l_2}} + \frac{8\gamma^2}{K_{l_2}}\int_0^1 \sigma_t^2 dt \\ &\quad + \left(1 + \frac{(K_{l_1}-1)(3K_{l_2}-K_{l_1}-1)}{3K_{l_2}}\right)\int_0^1 2\sigma_t^4 dt \\ &\quad \text{for } 1 \le l_1 < l_2 \le L, \end{aligned}\right\} \tag{20}$$

where $\gamma = 0$ under Condition RO(b').

Thanks to Proposition 8, the previous regression Eq. (11) still applies when K_l's are fixed integers (i.e., $\kappa = 0$) and the UV estimator still takes the form $\widehat{\langle X, X\rangle}_1$ as defined in (12). The only difference lies in asymptotic variances. By the same arguments as in Section 3.1, Proposition 8 implies that, when K_l's are fixed integers, the asymptotic covariance matrix of the error term $\mathbf{e}$ in (11) is no longer $\widetilde{\mathbf{Q}}/n^{1-\kappa}$ but becomes $\mathbf{Q}/n$, where $\mathbf{Q}$ is given in (19) and (20). Hence, the asymptotic variance of UV estimator in the case when K_l's are fixed integers is not given by (15) but instead given by

$$\frac{1}{n}\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\omega_{l_1}\omega_{l_2}q_{l_1,l_2}, \tag{21}$$

where ω_{l_i}, $i = 1, 2$, are given by (13) and q_{l_1,l_2} is given by (19) and (20).

The same reasoning as in Section 3.2 leads to the following central limit theorem.

Theorem 2. *Under the conditions of Proposition 8, stably in law,*

$$\sqrt{n}\left(\widehat{\langle X, X\rangle}_1 - \int_0^1 \sigma_t^2 dt\right) \to \sqrt{\mathcal{V}}N,$$

where N is a standard Normal random variable which is independent of the underlying σ-field and

$$\mathcal{V} := \sum_{l_1=1}^{L}\sum_{l_2=1}^{L} q_{l_1,l_2}\widetilde{\omega}_{l_1}\widetilde{\omega}_{l_2},$$

$$\text{where } \widetilde{\omega}_l := \frac{1}{L} - \frac{\frac{1}{L}\sum_{j=1}^{L}\frac{1}{K_j}\left(\frac{1}{K_l} - \frac{1}{L}\sum_{j=1}^{L}\frac{1}{K_j}\right)}{\sum_{j=1}^{L}\frac{1}{K_j^2} - L\left(\frac{1}{L}\sum_{j=1}^{L}\frac{1}{K_j}\right)^2}. \tag{22}$$

Remark 7. The discussion following Proposition 8 points out the difference between the asymptotic results of cases $\kappa > 0$ and $\kappa = 0$. It is clearly seen from Theorems 1 and 2 that UV estimators in the two cases have different asymptotic variances and convergence rates. In Section 3.5, we provide details for choosing the appropriate asymptotic framework (i.e., $\kappa > 0$ or $\kappa = 0$) to apply in a data-driven way.

We next construct a feasible central limit theorem. Recall that

$$\mathcal{Q} = \frac{2n}{3K^2}\sum_{k=0}^{K-1}[Y,Y,Y,Y]_1^{\mathcal{G}_K^{(k)}}$$

as defined in (17). We next give a consistent estimator for the asymptotic variance $\mathcal{V}$ in (22). By (19) and (20), we notice that $\mathcal{V}$ is a continuous tri-variate function evaluated at $(\gamma, \int_0^1\sigma_t^2dt, \int_0^1\sigma_t^4dt)$ and denote it as

$$\mathcal{V} := \psi\left(\gamma, \int_0^1\sigma_t^2dt, \int_0^1\sigma_t^4dt\right),$$

where the continuous function ψ is given as follows: with $\widetilde{\omega}_{l_i}, i = 1, 2$, being given in (22),

$$\psi(x,y,z) := \sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\widetilde{\omega}_{l_1}\widetilde{\omega}_{l_2}q_{l_1,l_2}(x,y,z),$$

for $x \geq 0, y > 0, z > 0,\ q_{l_2,l_1}(x,y,z) = q_{l_1,l_2}(x,y,z)$;

and for $1 \leq l \leq L$, $1 \leq l_1 < l_2 \leq L$, $q_{l,l}(x,y,z)$ and $q_{l_1,l_2}(x,y,z)$ are given by

$$\begin{cases} q_{l,l}(x,y,z) &= \dfrac{12x^4}{K_l^2} + \dfrac{8x^2}{K_l}y + \left(1 + \dfrac{(K_l-1)(2K_l-1)}{3K_l}\right)z \\ q_{l_1,l_2}(x,y,z) &= \dfrac{8x^4}{K_{l_1}K_{l_2}} + \dfrac{8x^2}{K_{l_2}}y \\ & \quad + \left(1 + \dfrac{(K_{l_1}-1)(3K_{l_2}-K_{l_1}-1)}{3K_{l_2}}\right)z \end{cases}.$$

An estimator for γ can be given as follows,

$$\widehat{\gamma} := \begin{cases} 0 & \text{under Condition RO(b}'\text{);} \\ \left(\dfrac{[Y,Y]_1^{\text{all}} - \widehat{\langle X,X\rangle}_1}{2}\right)^{1/2} & \text{under Condition RA(b}'\text{).} \end{cases}$$

Finally, an estimator for the asymptotic variance $\mathcal{V}$ is defined as follows,

$$\widehat{\mathcal{V}} := \psi\left(\widehat{\gamma}, \widehat{\langle X,X\rangle}_1, \frac{\mathcal{Q}}{2}\right). \tag{23}$$

We then have the following feasible central limit theorem.

Corollary 2. *Under the conditions of Theorem 2, if $K_L/K = o(1)$ under Condition RA(b') or $K_L/K = O(1)$ under Condition RO(b'), then stably in law,*

$$\sqrt{\frac{n}{\widehat{\mathcal{V}}}}\left(\widehat{\langle X,X\rangle}_1 - \int_0^1\sigma_t^2dt\right) \to Z,$$

where Z is a standard Normal random variable that is independent of the underlying σ-field, and $\widehat{\mathcal{V}}$ is as defined in (23).

3.5. Implementation of UV

In this section, we explain how the UV estimator used in our numerical studies is constructed. Two important ingredients of our scheme are given as follows first.

- **The intermediate situation.** To deal with the intermediate situation where $\gamma_n/\widetilde{\alpha_n}$ tends to some constant, we suggest further rounding the data to some coarser price grids α_n'. Notice that, for example, if the original n is used, $\sqrt{n}\widetilde{\alpha}_n'$ and/or $\sqrt{n}\gamma_n/\widetilde{\alpha}_n'$, where $\widetilde{\alpha}_n' := \alpha_n'/S_0$, may be too large to assume Condition RO(b'), but the (average) subsample grid size $\bar{n}_1 = o(n)$ can be selected such that $\sqrt{\bar{n}_1}\gamma_n/\widetilde{\alpha}_n'$ and $\sqrt{\bar{n}_1}\widetilde{\alpha}_n'$ have reasonable size to assume Condition RO(b). In effect, we take care of every situation.
- **Condition checking.** In practice, we let data tell the right theory to apply (Corollaries 1 and 2). We accomplish this in a unified fashion through checking Condition RO(b) or Condition RA(b). That is, let data tell the highest frequency (equivalently, the smallest K_1 or κ) at which Condition RO(b')/RO(b) or Condition RA(b')/RA(b) is met.

We propose a data-driven automated procedure. The procedure involves the choices of several parameters. For the choices of $K_1, \ldots, K_L$, we adopt a simple setting in which $K_{l+1} = K_1 + l$, $l = 1, 2, \ldots, L-1$. In the setup of (6), $(\mathcal{C}_l)_{1\leq l\leq L}$ is naturally determined as follows: $\mathcal{C}_1 = 1$, $\mathcal{C}_l = K_l/K_1 = 1 + (l-1)/K_1$ for $l = 2, \ldots, L$. There are still two parameters, K_1 and L, that need to be determined in the UV point estimator. For the standard error, one needs to decide another parameter K, and whether to use $\widetilde{\widetilde{\mathcal{V}}}$ defined in (18) or $\widehat{\mathcal{V}}$ defined in (23).

3.5.1. UV construction for a point estimation

We provide a data-adaptive procedure for choosing K_1 and hence all K_l's. As we discussed above, we should find a small K_1 through checking Condition RO(b) and Condition RA(b).[6] In the following, we explain the reasoning behind a simple and easy-to-implement procedure for practical use.

First, we try to understand the sizes of $\widetilde{\alpha_n}$ and γ_n. $\widetilde{\alpha_n}$ can be easily estimated by the observable counterpart $\widehat{\alpha_n} := \alpha_n/\breve{S}_0$. As to the unobservable γ_n, similar results to that of (4) and (5) under either Condition RA(b') or RO(b') lead to (cf. Remark 4)

$$\gamma_n^2 = \left([Y,Y]_1^{(\text{all})} - \langle X,X\rangle_1 - \text{LB}_{\text{ro}}(n) + \text{NLB}(n)\right)/(2n) + o_p(n^{-3/2}),$$

which provides a rough estimate for γ_n as follows

$$\widehat{\gamma_n} := \left\{\left([Y,Y]_1^{(\text{all})} - [Y,Y]_1^{(\text{avg},l_0)} - \widehat{\text{LB}_{\text{ro}}(n)} + \widehat{\text{NLB}(n)}\right)/(2n)\right\}^{1/2},$$

where

$$\widehat{\text{LB}_{\text{ro}}(n)} := \frac{1}{6}\frac{n\alpha_n^2}{\breve{S}_0^2} \quad \text{and}$$

$$\widehat{\text{NLB}(n)} := \frac{1}{\pi^2}\frac{n\alpha_n^2}{\breve{S}_0^2}\left\{\sum_{k=1}^{M}\frac{1}{k^2}\exp\left(-2\pi^2k^2\frac{[Y,Y]_1^{(\text{avg},l_0)}\breve{S}_0^2}{n\alpha_n^2}\right)\right\}, \tag{24}$$

are estimates of $\text{LB}_{\text{ro}}(n)$ (cf. (5)) and $\text{NLB}(n)$ (cf. (4)). Because the time horizon under consideration is usually short (e.g., one day), we use $\breve{S}_0$ to be a rough estimate of the level of S_t over $t \in [0,1]$, and $[Y,Y]_1^{(\text{avg},l_0)}$ at the 5-minute frequency (i.e., $K_{l_0} = 300$ s) to be rough estimates of the levels of both σ_t^2 over $t \in [0,1]$ and the target $\langle X,X\rangle_1$. The truncation of $\widehat{\text{NLB}(n)}$ made to the infinite series of $\text{NLB}(n)$ is justified by the uniform boundedness of the tail sum as follows:

$$\sum_{k=M+1}^{\infty}\frac{1}{k^2}\exp\left(-2\pi^2k^2\frac{\sigma_t^2S_t^2}{n\alpha_n^2}\right) < \int_M^{\infty}\frac{1}{x^2}dx = \frac{1}{M},$$

[6] Notice that we consider Condition RA(b) and Condition RO(b) directly to determine K_1 in point estimation. The construction of a point estimate is not affected by whether K_1 is viewed as finite. In calculation of standard error, we further determine whether K_1 should be viewed as finite in Section 3.5.2.

which can be arbitrarily small by letting M be large enough. We take $M = 1000$.

Second, given the estimates $\widehat{\alpha_n}$ and $\widehat{\gamma_n}$, we further check the rest of Condition RO(b) or Condition RA(b) to determine an appropriate K_1 to use in the simple linear regression. Condition RO(b) can be checked by evaluating the *nonlinear-to-linear bias ratio* due to rounding (at sampling frequency $\bar{n}_1$) defined as

$$\mathrm{NOL}(\bar{n}_1) := \mathrm{NLB}(\bar{n}_1)/\mathrm{LB}_{\mathrm{ro}}(\bar{n}_1).$$

In practice, if $\mathrm{NOL}(\bar{n}_1)$ is small (can be viewed as of order $o(1)$) then Condition RO(b) is fine. From our proof, it is clear that Condition RO(b) is required so that $\mathrm{NLB}(\bar{n}_1)$ can be negligible compared with $\mathrm{LB}_{\mathrm{ro}}(\bar{n}_1)$. Similarly, Condition RA(b) can be checked by evaluating the *random noise-to-signal ratio* (at sampling frequency $\bar{n}_1$) defined as

$$\mathrm{NSR}_{\mathrm{ra}}(\bar{n}_1) := \mathrm{LB}_{\mathrm{ra}}(\bar{n}_1)/\langle X, X\rangle_1,$$

where $\mathrm{LB}_{\mathrm{ra}}(\cdot)$ is as denoted in (5). If $\mathrm{NSR}_{\mathrm{ra}}(\bar{n}_1)$ is not large (can be viewed as of order $O(1)$), then Condition RA(b) is fine. The sample versions of $\mathrm{NOL}(\bar{n}_1)$ and $\mathrm{NSR}_{\mathrm{ra}}(\bar{n}_1)$ are given, respectively, as follows (cf. (5) and (24)):

$$\begin{aligned}\widehat{\mathrm{NOL}(\bar{n}_1)} &:= \frac{\widehat{\mathrm{NLB}(\bar{n}_1)}}{\widehat{\mathrm{LB}_{\mathrm{ro}}(\bar{n}_1)}}\\ &= \frac{6}{\pi^2}\sum_{k=1}^{M}\frac{1}{k^2}\exp\left(-2\pi^2k^2\frac{[Y,Y]_1^{(\mathrm{avg},l_0)}\breve{S}_0^2}{\bar{n}_1\alpha_n^2}\right) \text{ and}\\ \widehat{\mathrm{NSR}_{\mathrm{ra}}(\bar{n}_1)} &:= \frac{2\bar{n}_1\widehat{\gamma_n}^2}{[Y,Y]_1^{(\mathrm{avg},l_0)}}.\end{aligned}$$

The following gives an automated procedure for constructing the UV estimate.

Step 1. Determine whether the situation falls in Condition RA(b')/RA(b) or RO(b')/RO(b). In intermediate cases, conduct further rounding to obtain data that satisfies Condition RO(b). For a chosen $\theta > 1$,

- If $\widehat{\alpha_n}/\widehat{\gamma_n} \le 1/\theta$, then go to Step 2(RA);
- If $\widehat{\alpha_n}/\widehat{\gamma_n} \ge \theta$, then go to Step 2(RO);
- If $\widehat{\alpha_n}/\widehat{\gamma_n} \in (1/\theta, \theta)$, then let $\alpha_n' = (2\lfloor\lceil\widehat{\gamma_n}\breve{S}_0\theta/\alpha_n\rceil/2\rfloor + 1)\alpha_n$,[7] further round the data using the new tick size α_n', reset $\alpha_n = \alpha_n'$ and go to Step 2(RO).

We choose $\theta = 2$ in our numerical studies.

Step 2. Choose K_1.

- **Step 2(RA).** Find the smallest K_1 (i.e., the largest $\bar{n}_1$) among those with which the random noise-to-signal ratio $\widehat{\mathrm{NSR}_{\mathrm{ra}}(\bar{n}_1)}$ is less than ξ. Then go to Step 3. We choose $\xi = 1$ in our numerical studies.
- **Step 2(RO).** Find the smallest K_1 (i.e., the largest $\bar{n}_1$) among those with which the nonlinear-to-linear bias ratio due to rounding $\widehat{\mathrm{NOL}(\bar{n}_1)}$ is less than r_{nol}. Then go to Step 3. We implement $r_{\mathrm{nol}} = 0.01$ in our numerical studies.

Step 3. Obtain the final estimate.

- Run the simple linear regression based on data $\{\bar{n}_l, [Y,Y]_1^{(\mathrm{avg},l)}\}_{l=1}^L$ with $L = 10$ and K_1 given in Step 2. The intercept of the regression is taken as our final UV estimate of the integrated volatility.

Remark 8. We find in our unreported numerical studies that our approach is rather robust to the choices of $\theta \in [2,3]$, $\xi \in [1,2]$, $r_{\mathrm{nol}} \in [0.005, 0.01]$ and $L \in [5, 20]$.

3.5.2. Asymptotic variance estimation towards interval inference and beyond

Our feasible central limit theorems, Corollaries 1 and 2, allow one to conduct interval inference by further constructing asymptotic variance estimates $\widehat{\widetilde{\mathcal{V}}}$ defined in (18) and $\widehat{\mathcal{V}}$ defined in (23). For this, we need to determine whether K_1 selected from the UV construction procedure is finite or divergent. The following gives a guideline.

- We compare K_1 with K_c. If $K_1 < K_c$, adopt Corollary 2 and $\widehat{\mathcal{V}}$ defined in (23) to compute the asymptotic variance estimate; and adopt Corollary 1 and $\widehat{\widetilde{\mathcal{V}}}$ defined in (18) otherwise. We implement $K_c = 10$ in our numerical studies.[8]

As to the choice of K for estimating asymptotic variance, we take $K = [K_L n^{0.1}]$ under RA(b')/RA(b) (if Step 2(RA) in Section 3.5.1 is detected) and $K = 2K_L$ under RO(b')/RO(b) (if Step 2(RO) in Section 3.5.1 is detected) in our numerical studies.[9]

4. Monte-Carlo simulations

In this section, we investigate the performance of the proposed UV estimator in comparison with widely used methods through simulations. We include five widely used random-noise-oriented estimators and two rounding-error-oriented methods in our comparisons.

We refer readers to Zhang et al. (2005), Zhang (2006), Barndorff-Nielsen et al. (2008), Jacod et al. (2009) and Xiu (2010), respectively, for the definitions of five random-noise-oriented estimators. In particular, we choose parameter values of MSRV, RK, PAV and QMLE in an "optimal"[10] way as suggested in the corresponding papers which is summarized in Section 3.2 of Li et al. (2016). Because Li et al. (2016) does not provide details about TSRV, we illustrate this method and the choice of its "optimal" parameter value as follows. Under the pure additive noise assumption (see p. 1402 in Zhang et al. (2005)), the TSRV estimator is constructed as follows

$$\mathrm{TSRV} = \left(1 - \frac{n - K_{\mathrm{tsrv}} + 1}{nK_{\mathrm{tsrv}}}\right)^{-1} \times \left(\frac{1}{K_{\mathrm{tsrv}}}\sum_{k=0}^{K_{\mathrm{tsrv}}-1}[Y,Y]_1^{\mathcal{G}_{\mathrm{tsrv}}^{(k)}} - \frac{1}{K_{\mathrm{tsrv}}}[Y,Y]_1^{(\mathrm{all})}\right),$$

7 Further rounding using a new tick size α_n' over the rounded data that has old tick size α_n is equivalent to rounding using the same new tick size α_n' over the (unobservable) original data only if α_n' is odd multiples of α_n. In addition, the new tick size α_n' should be chosen as the smallest one such that $\alpha_n'/(\widehat{\gamma_n}\breve{S}_0) \ge \theta$, leading to the formula $\alpha_n' = (2\lfloor\lceil\widehat{\gamma_n}\breve{S}_0\theta/\alpha_n\rceil/2\rfloor + 1)\alpha_n$. For example, if $\alpha_n = 0.01$, then the new tick size α_n' must be chosen from $\{0.03, 0.05, 0.07, 0.09, \ldots\}$, depending on the values of $\widehat{\gamma_n}$, $\breve{S}_0$ and θ.

8 Numerical studies found that the results are rather robust to the choice of $K_c \in [1, 15]$.

9 One could also simply take $K = [K_L n^{0.1}]$ in all situations in practice. Mathematically, by writing $K = [K_L n^{0.1}]$ we are letting K be of a larger order than K_L, which is required under Condition RA(b')/RA(b), while $K = 2K_L$ means it is of the same order as K_L which is required under Condition RO(b')/RO(b). In fact, if K is of a larger order than K_L, consistency still holds under Condition RO(b')/RO(b). Numerically, for the sample sizes n considered in this paper, $[n^{0.05}, n^{0.15}]$ is about $[1.5, 6]$, and $[K_L n^{0.1}]$ is roughly similar to $2K_L$.

10 There are usually no real "optimal" choices of tuning parameters in these methods when the volatility process is time varying. See, e.g., Remark 3 on page 2256 of Jacod et al. (2009), Christensen et al. (2010) and Christensen et al. (2014). In our unreported simulation studies, we have also tried fixed tuning parameters for different observation frequencies and the conclusions are similar to those drawn from using the "optimal" ones, i.e., Tables 1 and 2.

Table 1
Simulation results for Design I: performance comparison among different methods based on 1000 replications. Values other than those of S_0 and CR in the table are reported as multiples of 10^{-6}. Feasible central limit theorem is not available for the method CPV of Rosenbaum (2009), hence, CR and CIwidth are not available for this estimator. When the rounding effect is weak ($S_0 = 50$), the UV method has a slight advantage over, or is comparable to, the random-noise-oriented methods in terms of bias and s.d. In addition, UV provides more accurate CR, implying that the central limit theorem of UV gives clearly better inference. In the intermediate cases ($S_0 = 20$ and 10), the random-noise-oriented methods are all severely biased, whereas the bias's of UV are substantially smaller. In these two cases, CR's of UV are much closer to the nominal level 95%. When the rounding effect is profound ($S_0 = 5$), the UV method substantially outperforms the random-noise-oriented methods in terms of bias, s.d. and CR. CPV is severely biased in all four cases. The method of Li and Mykland (2015) performs well when rounding is more severe but not as well when random noise dominates.

S_0	TSRV					MSRV					RK				
	bias	s.d.	rmse	CR	CIwidth	bias	s.d.	rmse	CR	CIwidth	bias	s.d.	rmse	CR	CIwidth
50	−0.009	4.1	4.0	82%	10.9	−3.4	4.0	5.2	81%	14.2	0.06	4.2	4.2	92%	13.8
20	11.5	5.2	12.7	22%	15.3	6.1	4.8	7.7	77%	18.8	7.2	5.0	8.8	67%	18.3
10	148.8	15.9	149.7	0%	34.6	131.6	13.9	132.3	0%	41.7	125.3	13.6	126.0	0%	40.3
5	470.3	61.1	474.3	0%	84.4	431.5	54.4	434.9	0%	101.3	421.7	51.1	424.8	0%	98.2

S_0	PAV					QMLE					UV				
	bias	s.d.	rmse	CR	CIwidth	bias	s.d.	rmse	CR	CIwidth	bias	s.d.	rmse	CR	CIwidth
50	−4.8	5.4	7.2	61%	14.1	−0.04	3.8	3.8	92%	13.4	−1.0	7.6	**7.7**	**96%**	32.7
20	16.1	6.5	17.3	16%	18.9	10.7	4.8	11.8	34%	17.8	−0.6	7.7	**7.8**	**96%**	33.5
10	119.7	13.0	120.4	0%	44.5	126.1	13.7	126.9	0%	40.3	1.7	14.5	**14.6**	**96%**	61.7
5	380.1	45.6	382.8	0%	123.2	407.6	77.4	414.9	0%	103.2	5.2	31.5	**31.9**	**93%**	115.0

S_0	LMV					CPV				
	bias	s.d.	rmse	CR	CIwidth	bias	s.d.	rmse	CR	CIwidth
50	239.1	34.6	241.6	0%	20.1	469.9	8.5	469.9	NA	NA
20	55.7	12.0	57.0	0%	22.3	473.8	19.0	474.2	NA	NA
10	14.3	16.2	21.6	54%	33.8	506.2	45.9	508.2	NA	NA
5	−3.8	31.4	31.6	61%	54.0	653.5	122.4	664.9	NA	NA

Table 2
Simulation results for Design II: performance comparison among different methods based on 1000 replications. Values other than those of S_0 and CR in the table are reported as multiples of 10^{-6}. Feasible central limit theorem is not available for the method CPV of Rosenbaum (2009), hence, CR is not available for this estimator. When the rounding effect is weak ($S_0 = 50$), the UV method has a slight advantage over, or is comparable to, the random-noise-oriented methods in terms of bias and rmse. In addition, UV provides more accurate CR, implying that the central limit theorem of UV gives clearly better inference. In the intermediate cases ($S_0 = 20$ and 10), the random-noise-oriented methods are all severely biased, whereas the bias's of UV are substantially smaller. In these two cases, CR's of UV are much closer to the nominal level 95%. When the rounding effect is profound ($S_0 = 5$), the UV method substantially outperforms the random-noise-oriented methods in terms of bias, rmse and CR. CPV is severely biased in all four cases. The method of Li and Mykland (2015) performs well when rounding is more severe but not as well when random noise dominates.

S_0	TSRV			MSRV			RK			PAV			QMLE			UV		
	bias	rmse	CR	bias	rmse	CR	bias	rmse	CR	bias	rmse	CR	bias	rmse	CR	bias	rmse	CR
50	−0.2	3.9	83%	−3.6	5.3	82%	−0.09	4.0	92%	−2.2	7.6	60%	−0.2	3.8	92%	−1.2	**7.2**	**97%**
20	11.9	13.2	22%	6.3	8.1	73%	7.6	9.2	62%	13.6	15.5	28%	10.9	12.1	35%	−0.2	**7.4**	**97%**
10	146.4	147.7	0%	129.8	130.8	0%	123.9	125.0	0%	118.1	119.1	0%	125.0	126.5	0%	1.6	**14.5**	**96%**
5	469.3	475.0	0%	429.8	434.7	0%	418.3	422.9	0%	379.0	383.2	0%	406.2	416.0	0%	6.7	**32.6**	**93%**

S_0	LMV			CPV		
	bias	rmse	CR	bias	rmse	CR
50	252.8	261.0	0%	469.5	469.5	NA
20	56.9	58.5	0%	474.4	474.7	NA
10	12.9	20.9	57%	505.8	508.1	NA
5	−3.2	30.5	62%	655.5	668.3	NA

where $K_{\text{tsrv}} = [c_{\text{tsrv}} n^{2/3}]$ and $[Y,Y]_1^{\mathcal{G}_{\text{tsrv}}^{(k)}} := \sum_{t_i \in \mathcal{G}_{\text{tsrv}}^{(k)}} (Y_{t_{i,+}} - Y_{t_i})^2$ with $\mathcal{G}_{\text{tsrv}}^{(k)} := \{t_{k+iK_{\text{tsrv}}} \in \mathcal{G} : i = 0, 1, 2, \ldots\}$ for $k = 0, 1, \ldots, K_{\text{tsrv}} - 1$, $\mathcal{G}$ being the full grid of times, and $t_{i,+}$ being the time that follows t_i in $\mathcal{G}_{\text{tsrv}}^{(k)}$. We adopt the "optimal" constant

$$c_{\text{tsrv}} = \left(\frac{12(\widehat{E\varepsilon^2})^2}{\widehat{Q_1^n}} \right)^{1/3},$$

which minimizes the expected asymptotic variance of TSRV when $(\sigma_t)_{0 \le t \le 1}$ is constant (cf. Eq. (63) in Zhang et al., 2005), where $\widehat{Q_1^n}$ is given by (3.14) in Jacod et al. (2009) and $\widehat{E\varepsilon^2} := [Y,Y]_1^{(\text{all})}/(2n)$ is an estimate of the variance of the noise. The asymptotic variance can be estimated by (p.1403, Example 1 in Zhang et al., 2005)

$$n^{-1/3} \left\{ 8c_{\text{tsrv}}^{-2} (\widehat{E\varepsilon^2})^2 + \frac{4}{3} c_{\text{tsrv}} \widehat{Q_1^n} \right\}.$$

Next, we introduce a bit more about the rounding-error-oriented methods used in our numerical studies. The first one with which we compare is the so called *compensated estimator* of Rosenbaum (2009, Section 2.4 on p. 691). We abbreviate this method as CPV in our comparison. The second rounding-error-oriented method is that of Li and Mykland (2015) which is abbreviated as LMV in our studies. In order to meet the required conditions of Li and Mykland (2015), LMV involves the choice of an appropriate subsampling frequency. For fairness of comparison, we select this sampling frequency by using Step 2(RO) of the automated procedure of UV.

We make the comparisons under two simulation designs. Condition RO(c), which requires that the volatility is a function of the price, is satisfied in the first simulation design and is violated in the second.

4.1. Design I: The Ornstein–Uhlenbeck model

We first consider the Ornstein–Uhlenbeck model for the latent log price process:

$$dX_t = \vartheta(\nu - X_t)dt + \sigma dW_t. \tag{25}$$

We set $\vartheta = 5/252$ and $\sigma = 0.0125$. The parameter ν gives the long-term mean of the mean-reverting process (25). In our simulation, we set the value of ν to be the same as the initial value of the process $(X_t)_{0\le t\le 1}$.

Comparisons are made under four different settings for (S_0, ν): $(50, \log(50))$, $(20, \log(20))$, $(10, \log(10))$ and $(5, \log(5))$, with fixed $\alpha_n = 0.01$ and $\gamma_n = 10^{-4}$. These settings represent situations of weak rounding to strong rounding ($\widetilde{\alpha_n} = 0.01/50$ to $\widetilde{\alpha_n} = 0.01/5$). We consider a time horizon of one trading day and simulate one observation per second, giving a total number of 23,401 observations for each sample path.

Table 1 provides the finite sample performance comparison between the alternative methods and our method over 1000 simulated sample paths. We report the finite sample bias (`bias`), standard deviation (`s.d.`), root mean squared error (`rmse`), the coverage rate (`CR`) and average width (`CIwidth`) of the 95% confidence interval over 1000 replications. Overall, the performance of our method is superior over that of the alternative methods and stable across different rounding level settings. Specific findings are summarized as follows.

- The weak rounding case ($S_0 = 50$) is very challenging for UV as the rounding effect is almost negligible, and hence the random-noise-oriented estimators may have advantages because they are designed to work when rounding can be ignored. However, even in this case, the performance of UV still shows a slight advantage over, or is comparable to, that of the random-noise-oriented estimators in finite sample bias and standard deviation. In addition, the confidence interval of UV provides more accurate coverage rate, implying that the central limit theorem of UV gives clearly better inference.
- In the intermediate situations ($S_0 = 20$ and 10), the random-noise-oriented methods are all severely biased, whereas the finite sample bias of our approach is substantially smaller. In these two cases, the confidence interval of UV provides much more accurate coverage rates that are clearly closer to the nominal level 95%.
- In the strong rounding case ($S_0 = 5$), our method has substantial bias reduction and also superior performance in finite sample standard deviation compared with the random-noise-oriented methods. Once again, the confidence interval of UV provides coverage rate that is close to the nominal level 95% in this case. In sharp contrast, the random-noise-oriented estimators have coverage rates close to 0%.
- The method CPV of Rosenbaum (2009) is severely biased in all four cases considered in this simulation design. This can be understood[11] because this rounding-error-oriented method is designed for dealing with pure rounding case.
- The performance of the method LMV of Li and Mykland (2015) deteriorates as rounding effect becomes smaller. LMV provides less accurate coverage rates than UV in all four cases considered in this simulation design.

In summary, random-noise-oriented methods (rounding-error-oriented methods, respectively) may work fine when rounding is negligible (when rounding dominates, respectively). UV is the only method that consistently generates good point and interval estimates in all situations.

The histograms and Q–Q plots shown in Fig. 3 further confirm our central limit theorems developed in Section 3.

[11] The implementation of CPV involves the choice of a tuning parameter $0 < a < 1$ (Rosenbaum, 2009, Section 2.4 on p. 691). We did not find a guideline on choosing this parameter in Rosenbaum (2009). In our numerical studies, we find that the results are similar for $a \in [0.1, 0.9]$ and only report the results of $a = 0.5$. Note also that there is no feasible central limit theorem in Rosenbaum (2009) so we left the interval estimates blank.

4.2. Design II: The Heston model

We also consider the Heston stochastic volatility model for the latent log price process $(X_t)_{0\le t\le 1}$:

$$\begin{cases} dX_t = \left(\mu - \dfrac{V_t}{2}\right)dt + \sqrt{V_t}dW_t \\ dV_t = \vartheta_V(\nu_V - V_t)dt + \upsilon\sqrt{V_t}dW_t^\sigma \end{cases},$$

where parameters μ, ϑ_V, ν_V and υ are positive constants and $(W_t)_{0\le t\le 1}$ and $(W_t^\sigma)_{0\le t\le 1}$ are standard Brownian motions with instantaneous correlation coefficient ρ. We set $\mu = 0.05/252$, $\vartheta_V = 5/252$, $\nu_V = 0.0125^2$, $\upsilon = 0.5/252$ and $\rho = -0.5$. We take the simulation studies in Zhang et al. (2005) as reference when choosing the parameter values. Again, the noise setting $\alpha_n = 0.01$ and $\gamma_n = 10^{-4}$ is used and comparison is made under four different price levels $S_0 = 50,\ 20,\ 10$ and 5. We consider a time horizon of one trading day and simulate one observation per second, giving a total number of 23,401 observations for each sample path.

Notice that this is a setup where Condition RO(c) is violated. Nevertheless, our method is still seen to have its clear advantages. The aim of this paper is to demonstrate the idea of such a unified approach for volatility estimation. A formal technical extension of our results to relax Condition RO(c) requires a generalization of the results of Delattre and Jacod (1997) to allow volatility to be general stochastic processes.

Table 2 again confirms the superior performance of our method over that of the alternative methods with the violation of Condition RO(c). The findings based on 1000 replications are similar to those in Section 4.1. More specifically,

- In the weak rounding case ($S_0 = 50$), the performance of UV still shows a slight advantage over, or is comparable to, that of the random-noise-oriented estimators in finite sample bias. In addition, the confidence interval of UV provides more accurate coverage rate, implying that the central limit theorem of UV gives clearly better inference.
- UV has substantially smaller finite sample bias and root mean squared error than the random-noise-oriented methods in the cases when rounding effect is not negligible ($S_0 = 20,\ 10,\ 5$). In these cases, the confidence interval of UV provides much more accurate coverage rates that are clearly closer to the nominal level 95%. In sharp contrast, the random-noise-oriented estimators even have coverage rates close to 0% in the latter two cases ($S_0 = 10,\ 5$).
- The method CPV of Rosenbaum (2009) is severely biased in all four cases considered in this simulation design. The bias of the method LMV of Li and Mykland (2015) increases when random noise gets dominant (as S_0 goes from 5 to 50). When rounding effect is strong ($S_0 = 5$), LMV and UV produce similar volatility estimates. However, UV provides better inference than LMV as can be clearly seen from the coverage rates of their confidence intervals.

In summary, we reach similar conclusions as before. UV is the only method that consistently generates good point and interval estimates in all situations.

5. Empirical studies

We apply UV to both stock and currency exchange data and compare UV with the alternative methods. Stocks have a rounding tick size of 0.01 and exchange rates considered in this paper have smaller rounding tick sizes ranging from 10^{-5} to 10^{-3}. The empirical findings also show clear advantages of UV in practice.

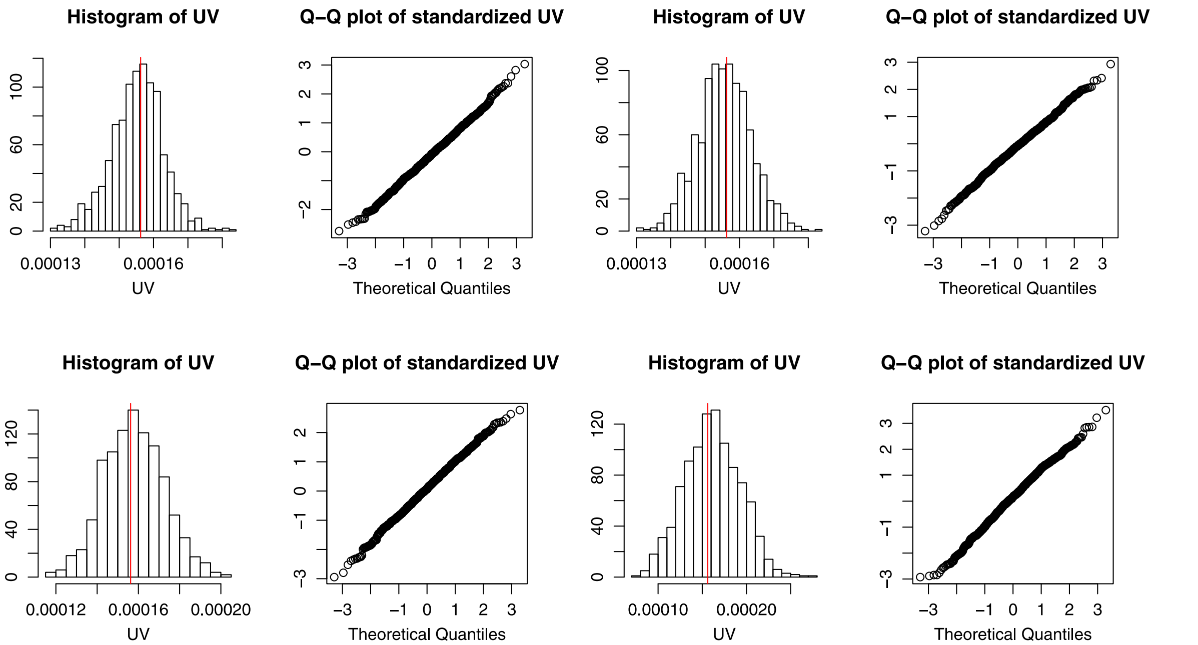

Fig. 3. Histograms for the UV estimates and Q–Q plots for the standardized UV estimates based on 1000 replications under Design I. Standardization is done according to Section 3.5.2 (Corollary 1 or Corollary 2). Placed from left to right and top to bottom are cases $S_0 = 50,\ 20,\ 10$ and 5, respectively.

5.1. Stocks data and performance of UV

As mentioned in the Introduction, we consider four stocks, Citigroup Inc. (NYSE: C), Intel Corporation (NASDAQ: INTC), Bank of America (NYSE: BAC) and Microsoft Corporation (NASDAQ: MSFT), which are all actively traded. The high-frequency trade data can be obtained from the NYSE Trade and Quote (TAQ) database of the Wharton Research Data Services. We randomly choose a date (May 1, 2009, with time span from 9:30:00 to 16:00:00) to study these stocks. The numbers of observations of the original tick-by-tick data sets are 119,228, 144,993, 396,695 and 167,962, respectively, for C, INTC, BAC and MSFT. The one-second-data sets are constructed by setting the close price of every one-second interval as the price for the corresponding second. The previous-tick (Zhang, 2006) method is then used to construct the second-by-second data whenever there is a value missing on a particular time stamp. Hence, for each stock, we have 23,401 second-by-second observations in total. We build equally spaced data because regular spacing is assumed by RK, PAV, QMLE, CPV and UV in estimating the integrated volatility of the logarithm of the latent efficient price process. Furthermore, we apply a Zhou (1996) type validation procedure to the second-by-second data to remove outliers due to, e.g., keying errors.[12] The resulting second-by-second price paths are shown in Fig. 2 in the Introduction. The rounding level α_n is 0.01 for the four selected stocks.

We examine the "volatility signature plot" (Andersen et al., 2000) to check the performance of UV in comparison with the alternative methods in practice. For each stock, we apply each method to price data with different sampling intervals. In Fig. 4, circles overplotted with solid lines indicate those estimates provided by our UV method. Because these estimates are for the same stock over the same day, a good volatility estimator is expected to generate similar estimates over different sampling intervals.

Specific findings from Fig. 4 are listed as follows.

- The estimates provided by the methods under comparison all diverge when the sampling interval becomes smaller (sampling frequency gets larger), exhibiting a volatility signature plot pattern.
- The UV method remains stable across different sampling intervals.

Through a close examination of the intermediate outputs, we find that our UV procedure selects K_1 through (RO) in Step 2 for all the four cases, which means rounding dominates in this study. Then take MSRV in the experiments with stock INTC for example, its volatility signature plot pattern is related to the selection of its tuning parameter (i.e., θ on page 40 of Li et al., 2016). The selected θ for MSRV at the five sampling intervals are 0.02, 0.028, 0.035, 0.04 and 0.047, which change moderately and lead to the effective subsample frequencies (n_s) 5850, 2925, 1950, 1463 and 1170. This explains the signature plot pattern since the (linear) bias due to rounding is in the form of the effective subsample frequency multiplied by a factor depending on the error level. By contrast, the corresponding effective subsample frequencies selected by our UV procedure are 4680, 3900, 3900, 2925 and 4680, which do not exhibit apparent trend, yielding stable (unbiased) estimates. The results of empirical experiments for the other random-noise-oriented methods and stocks in this regard are similar. On the other hand, random noise is still non-negligible even though in these cases rounding dominates, therefore, the biases exhibit in the rounding-error-oriented estimators CPV and LMV.

5.2. The effect of further rounding

The empirical analysis above does not involve further rounding. In this section, we discuss the issue of further rounding that might occur when applying our procedure. We intentionally modify the

[12] Zhou (1996) validates an exchange rate quote by comparing it to the medians of three neighboring points on both sides. We apply their procedure to second-by-second trade data of stocks here. A price other than those on the initial and last three time stamps is removed when it is $10 * \alpha_n$ above or below the medians of three neighboring points on both sides. The median of three neighboring points that follow (resp. precede) each of the initial (resp. last) three prices is used in validating each of the initial (resp. last) three prices. Whenever a price on a particular time stamp is removed as an outlier, it is replaced by the price on the previous time stamp except that if an outlier is identified on the initial three time stamps then it is replaced by the median of three neighboring points that follow it. The numbers of identified outliers for the four selected stocks C, INTC, BAC and MSFT are, respectively, 0, 0, 1 and 3 out of 23,401 observations.

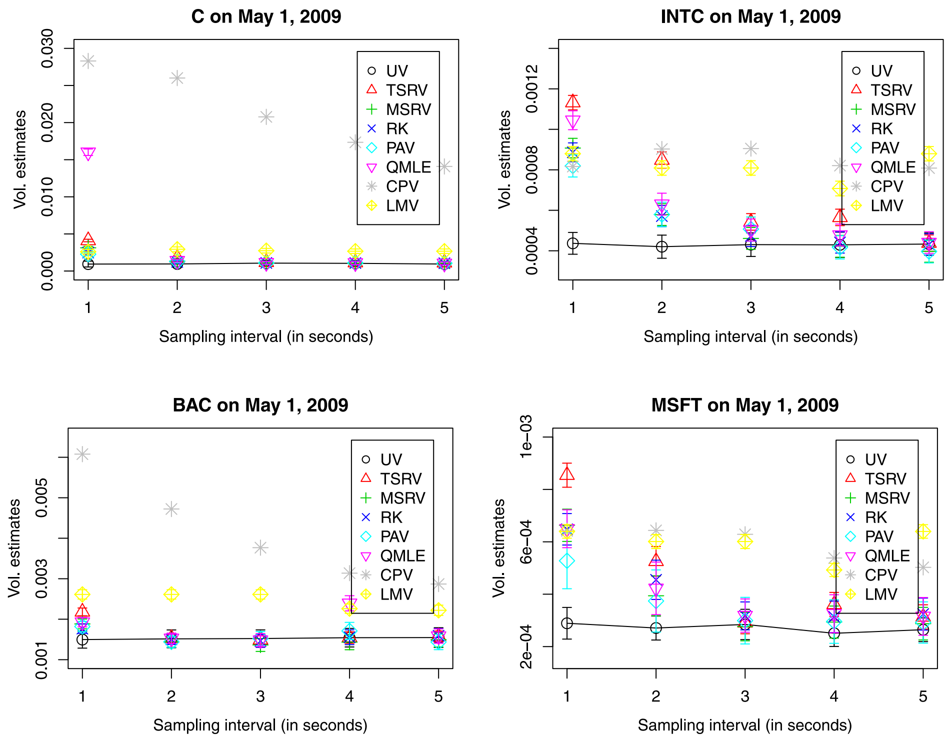

Fig. 4. The UV estimates and the volatility estimates of methods under comparison across different sampling intervals based on the second-by-second data of four stocks on May 1, 2009. The vertical segment lines indicate the 95% confidence intervals of the corresponding estimators. The most stable estimates are seen with the UV method (circles connected with solid lines) across different sampling intervals in the comparison. The estimates given by the alternative methods all diverge as the sampling interval becomes smaller.

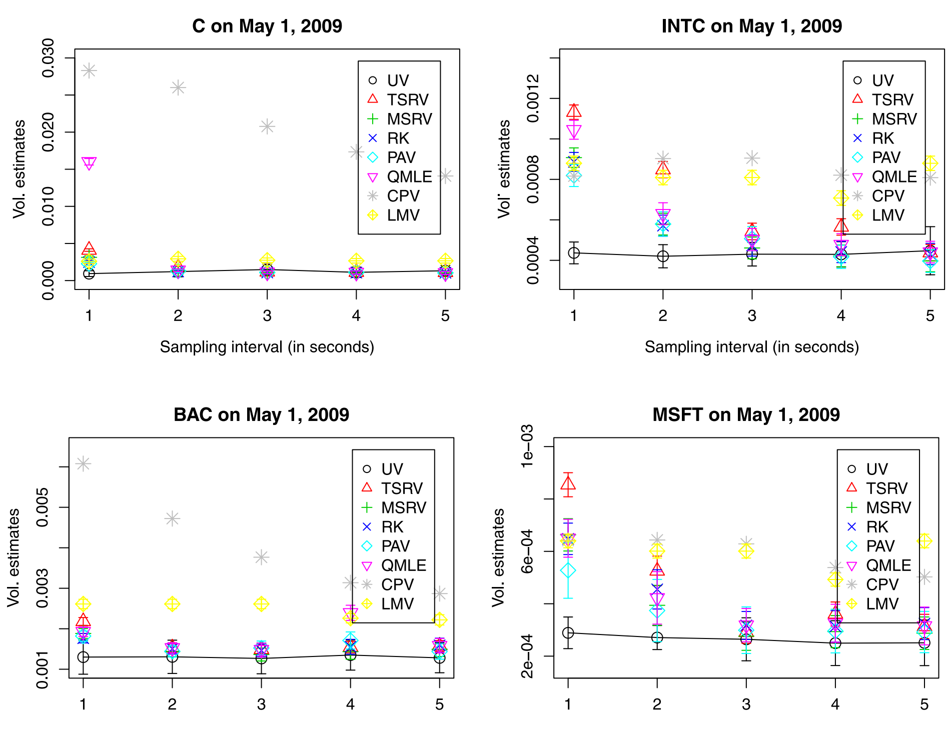

Fig. 5. The UV estimates with further rounding and the volatility estimates of methods under comparison across different sampling intervals based on the second-by-second data of four stocks on May 1, 2009. The UV estimates are produced by a modified UV procedure of Section 3.5.1 with θ being changed from 2 to 3. When comparing Fig. 5 to Fig. 4, one sees that the difference in performance of our method between the two cases is slight.

UV procedure given in Section 3.5.1 by letting $\theta = 3$ instead of $\theta = 2$ and apply the modified procedure to the same stock data as in Section 5.1. Fig. 5 reports the estimation results together with a comparison to the alternative methods. In this case, all stocks incur further rounding with the rounding levels of C, INTC, BAC and MSFT changed from the previous 0.01 to 0.03. When comparing Fig. 5 to

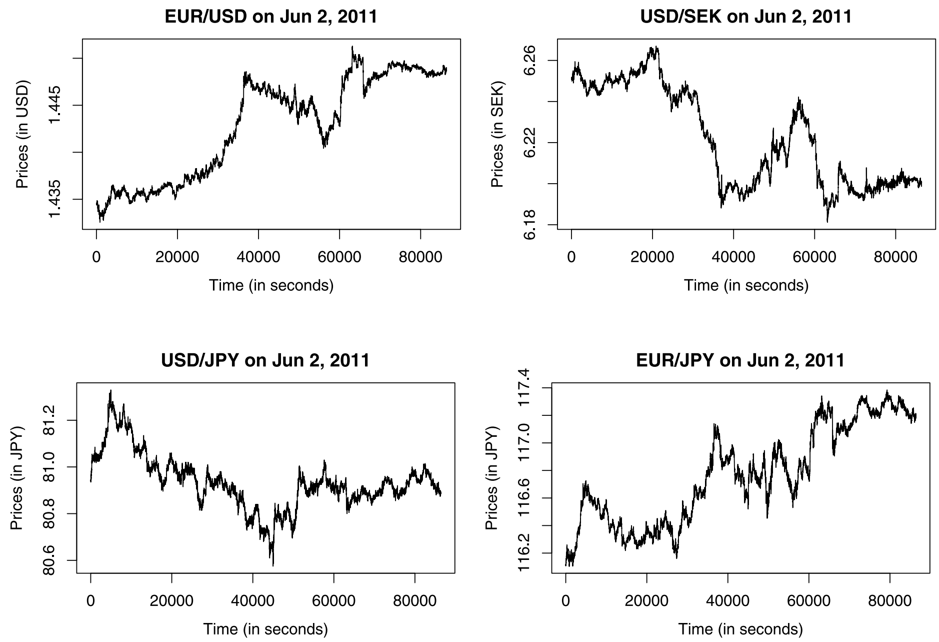

Fig. 6. Second-by-second mid-quotes of four currency pairs on June 2, 2011.

Fig. 4, one sees that the difference in performance of our method between the two cases is small.

Our findings based on this study are as follows.

- Estimation results are rather robust to the choice of tuning parameter θ in Section 3.5.1.
- When indeed an intermediate case (i.e., when neither Condition RA(b')/RA(b) nor Condition RO(b')/RO(b) holds) is detected, applying further rounding before using our theory can produce good results.

We emphasize that this subsection is for illustration only. In practice, we still encourage the use of a moderate-sized θ as suggested in Section 3.5.1 to avoid additional error introduced by too much further rounding.

5.3. Beyond assumed settings

5.3.1. Exchange rates data

As a common practice in foreign exchange markets, currency pairs are often quoted in multiples of the percentage in points (pips) from 10^{-3} to 10^{-6}, depending on price levels. Hence, it is interesting to investigate the performance of different methods when applied to exchange rate data. Unlike stock markets, foreign exchange markets have no geographical location and no business-hour limitations. Exchange rates are traded over-the-counter and 24 h a day except weekends from 22:00 Friday to 22:00 Sunday. Hence, we can work with 86,401 rather than 23,401 observations per trading day for second-by-second data. Another distinction between foreign exchange markets and stock markets is that transaction prices are not known to the public in foreign exchange markets. In the following study, mid-quotes are used because they are thought to be closest to the true transaction prices (see Zhou, 1996).

We select four major currency pairs, EUR/USD, USD/SEK, USD/JPY and EUR/JPY, which are actively traded in the market. The one-second-data sets were provided by Olsen Financial Technologies. The data sets are constructed by setting the close quote for every one-second interval as the price for the corresponding second. We randomly choose to study the data on June 2, 2011, which is a Thursday. Again, the previous-tick interpolation is used for the time stamps with missing values. Therefore, in the following study, we have 86,401 observations for each of the four data sets, which amounts to 24 trading hours a day and one trade per second. The second-by-second price paths are shown in Fig. 6. Roundings are made at 10^{-5}, 10^{-5}, 10^{-4} and 10^{-3}, respectively, for EUR/USD, USD/SEK, USD/JPY and EUR/JPY.

The estimates of different methods across different sampling intervals ("volatility signature plot") are shown in Fig. 7. Findings are:

- Again, UV clearly provides the most stable performance showing *no* signature plot pattern.
- The methods under comparison all exhibit a volatility signature plot pattern.

5.3.2. Tick-by-tick stocks data

We also investigate the performance of all methods when applied directly to the original tick-by-tick stock data. That is, if one simply ignores the fact that the original tick-by-tick data is non-equally spaced, and applies the estimators directly. Notice again that regular spacing of times is assumed by RK, PAV, QMLE, CPV and UV. Hence, this is a situation where the required regular spacing assumption is violated for these estimators.

In Fig. 8, for each stock, we compare UV with the alternative methods based on 5 data sets sampled from the original tick-by-tick data with 5 different sampling frequencies. The data set with sampling step size of i $(= 1, 2, \ldots, 5)$ refers to the data with observation time grid $\{t_{qi} \in \mathcal{G} : q = 0, 1, 2, \ldots\}$, where $\mathcal{G}$ denotes the full grid of times corresponding to the original tick-by-tick data. The practice of using data sampled in tick time for empirical analysis can be found in, e.g., Barndorff-Nielsen et al. (2008).

Findings from Fig. 8 are:

- The estimates given by the alternative methods except for LMV all diverge as the sampling step size approaches 1, exhibiting volatility signature plot patten. The UV method keeps being stable across different sampling step sizes.

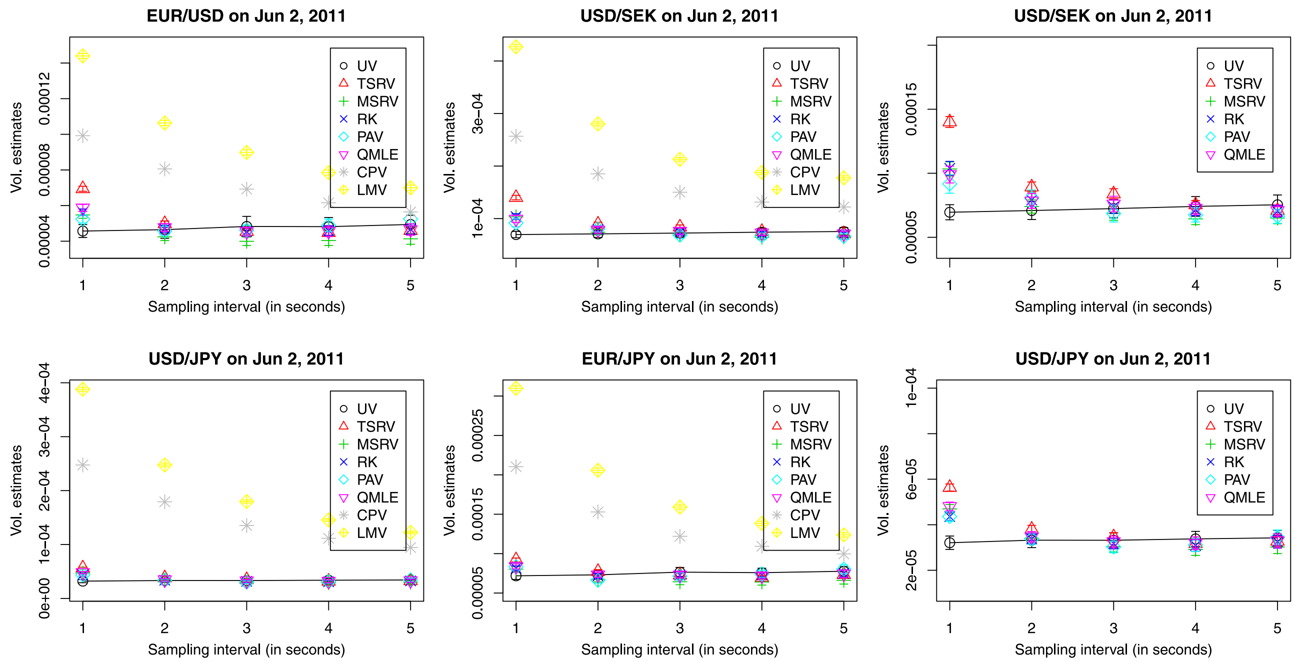

Fig. 7. The UV estimates and the volatility estimates of methods under comparison across different sampling intervals based on the second-by-second data of four currency pairs on June 2, 2011. The vertical segment lines indicate the 95% confidence intervals of the corresponding estimators. The most stable estimates are seen with the UV estimator (circles connected with solid lines) across different sampling intervals in the comparison. The estimates given by the alternative methods all diverge as the sampling interval becomes smaller. The two graphs on the right hand side are the zoomed-in versions of the estimation results for USD/SEK and USD/JPY, where LMV and CPV are removed to show more clearly the signature plots of the random-noise-oriented methods.

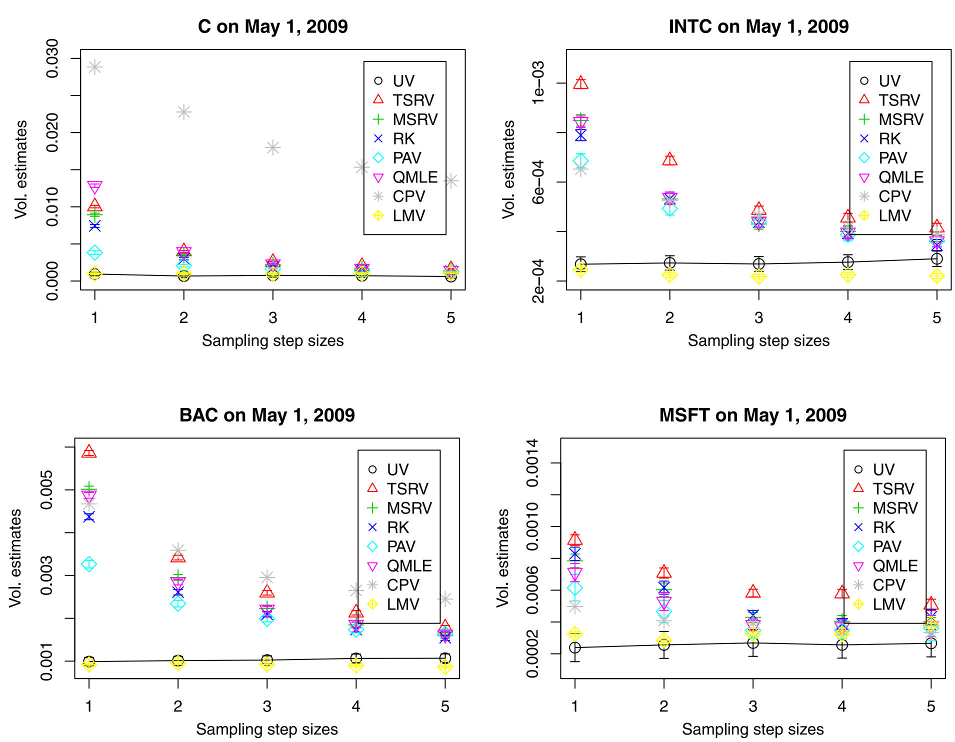

Fig. 8. The UV estimates and the volatility estimates of methods under comparison across different sampling step sizes based on the original tick-by-tick data of four stocks on May 1, 2009, ignoring the fact that the data is non-equally spaced. The vertical segment lines indicate the 95% confidence intervals of the corresponding estimators. The most stable estimates are seen with the UV method (circles connected with solid lines) across different sampling step sizes in the comparison. The estimates given by the alternative methods except for LMV all diverge as the sampling step size approaches 1.

- The volatility signature plot pattern of the random-noise-oriented methods for the tick-by-tick data is more severe than that for the second-by-second data, and the performance of LMV is the closest to that of UV. One reason for this is that the dominance of rounding over random noise in the tick-by-tick data is stronger than that in the second-by-second data. Hence, in this study, LMV effectively selects a subsampling frequency in the same way as UV (using (RO) in Step 2 of the UV procedure), yielding similar volatility estimates.

5.3.3. Discussions

Notice that the exchange rates data we discussed in Section 5.3.1 is an interesting scenario where rounding is not a main component of noise but UV again shows clear advantage. In fact, the rounding levels range from 10^{-5} to 10^{-3} for these exchange rates data. Our data driven procedure does further confirm that Condition RA, i.e., random noise dominates, is identified for all these four cases. Both rounding-error-oriented estimators CPV and LMV exhibit large biases as we can expect. After removing CPV and LMV from our comparison, we see clearly that our UV estimator still outperforms the random-noise-oriented estimators. In our unreported further empirical analyses, we found that this advantage is related to the positive serial dependence of random noise in the data, as documented in Jacod et al. (2017a). Moreover, as can be seen from Jacod et al. (2017b), the bias induced by serial dependence of noise in RV is also linear in sampling frequency, which explains the robustness of UV to dependent noise.

The study in Section 5.3.2 on the tick-by-tick stocks data is again highly encouraging. It shows that UV further enjoys advantages on data exhibiting additional features like irregular (may be endogenous) sampling times.

To avoid diluting the focus of the paper, we do not intend to establish theoretical results for the cases with further dependence in random noise and irregular sampling times. Potential advantages of UV may go well beyond scenarios discussed in this paper.

6. Concluding remarks

In this paper, we develop a unified approach to volatility estimation in the presence of both rounding and random microstructure noise. The proposed UV method has superior properties in terms of consistency and efficiency. Simulation and empirical studies further demonstrate significant advantages of our proposed method over the widely used volatility estimators. To be more specific,

- Our unified volatility estimator is the first method that achieves consistency under a wide spectral of situations where rounding and general additive noise co-exist.
- We find a common feature shared by the biases in RV caused by the two types of noise which are of distinct nature. This innovative observation leads to our unified volatility estimator in this paper, and should be useful for many more broadly related studies.
- We provide a data-adaptive automated procedure for selecting parameter values in our estimator such that each individual price process uses a set of parameter values that are tailor-made to itself towards an unbiased efficient volatility estimate.
- Simulation studies show that UV is the only estimator that consistently works well for a range of cases where random noise and rounding co-exist. In contrast, the random-noise-oriented or rounding-error-oriented estimators under comparison may work fine only when the other type of noise can be roughly negligible.
- Empirical studies on both stocks and exchange rates data show that UV keeps being stable across different sampling frequencies while the methods under comparison often exhibit apparent biases at high frequencies.

Our key findings about the linkage between rounding and random noise, and our methodology to robustly remove the impact of both should have broad implications in high-frequency data and market microstructure research.

Acknowledgments

We thank the Editor and two anonymous referees for their helpful comments and constructive suggestions. Yingying Li's research is supported in part by RGC grants GRF 602710, 16502014 and 16518716 of the HKSAR; Zhiyuan Zhang's research is supported by the National Natural Science Foundation of China (71301097) and the State Key Program of National Natural Science Foundation of China (91546202).

Appendix. Proofs

A.1. Conventions

The following convention is adopted throughout the proofs:

C, C_1 and C_2 are generic positive constants that may change from line to line.

We use $[X^{(i)}, X^{(j)}]_1^{\mathcal{G}_s} := \sum_{q=0}^{n_s-1} \Delta X_{\tau_q}^{(i)} \Delta X_{\tau_q}^{(j)},\ i, j = 1, 2,$ to denote realized (co-)volatilities of two generic processes $(X_t^{(i)})_{t\geq 0}$ and $(X_t^{(j)})_{t\geq 0}$ over the sub-grid $\mathcal{G}_s$ of times. The first order difference operator Δ over the sub-grid $\mathcal{G}_s$ always operates as $\Delta X_{\tau_q}^{(i)} = X_{\tau_{q+1}}^{(i)} - X_{\tau_q}^{(i)},\ i = 1, 2$. Similarly, we use $[X^{(i)}, X^{(j)}]_1^{\mathcal{G}_{k,l}} := \sum_{t_q \in \mathcal{G}_{k,l}} (X_{t_{q,+}}^{(i)} - X_{t_q}^{(i)})(X_{t_{q,+}}^{(j)} - X_{t_q}^{(j)}),\ i, j = 1, 2,$ to denote realized (co-)volatilities of two generic processes $(X_t^{(i)})_{t\geq 0}$ and $(X_t^{(j)})_{t\geq 0}$ over the sub-grid $\mathcal{G}_{k,l}$, where if $t_q \in \mathcal{G}_{k,l}$, then $t_{q,+}$ denotes the next time of t_q in $\mathcal{G}_{k,l}$.

Because we shall establish stable convergence in law and convergence in probability and because of the local boundedness condition on $(\mu_t)_{t\geq 0}$ and $(\sigma_t^2)_{t\geq 0}$, by standard localization arguments we can in the sequel assume the following without loss of generality:

1. $\mu_t \equiv 0$; and
2. $0 < \sigma_- \leq \sigma_t \leq \sigma_+$ for all $t \in [0, 1]$, where σ_- and σ_+ are nonrandom numbers.

See e.g. Mykland and Zhang (2009) and Mykland and Zhang (2012).

Let $M > 0$ and $m < 0$ be two constants. In view of Delattre and Jacod (1997), it suffices to prove results on $[0, \mathcal{T}]$, where $\mathcal{T} := \inf\{t : X_t \in (-\infty, m) \cup (M, \infty)\}$ and $\mathcal{T} \to \infty$ a.s. as $\min(|m|, M) \to \infty$. Hence, without loss of generality, we assume $X_t \in [m, M]$ for all $t \in [0, 1]$. Hence $S_t \in [\exp(m), \exp(M)]$ for all $t \in [0, 1]$. Since S_t is uniformly bounded for $t \in [0, 1]$, we can substitute α_n for $\widetilde{\alpha}_n$ in Condition RA(a)-(b)/RA(b') and Condition RO(a)-(b)/RO(b'). Therefore, the following proofs will be given under the Condition RA(a)-(b)/RA(b') and Condition RO(a)-(b)/RO(b') with $\widetilde{\alpha}_n$ being replaced by α_n.

A.2. Proofs for the results in Section 2.2

We first introduce some new notations that will be used later. Denote

$$\widetilde{S}_{\tau_i} := \left(e^{X_{\tau_i}}\right)^{(\alpha_n)},\ \widetilde{Y}_{\tau_i} := \log\left(\widetilde{S}_{\tau_i}\right) \text{ and } R_{\tau_i} := Y_{\tau_i} - \widetilde{Y}_{\tau_i} = \log\left(\frac{\check{S}_{\tau_i}}{\widetilde{S}_{\tau_i}}\right),$$

Hence, $Y_{\tau_i} = \widetilde{Y}_{\tau_i} + R_{\tau_i}$. Define

$$b_n := \alpha_n/\gamma_n. \tag{26}$$

Hence, when rounding effect dominates, $b_n \to \infty$. Let c_n be any auxiliary sequence such that $c_n \to \infty$ and $c_n/b_n \to 0$.

As long as n is large enough, all subsequent $\alpha_n < 2\exp(m)/3$ and $c_n/b_n < 1/2$, and hence $c_n\gamma_n < \alpha_n/2$. We consider only those large n's. Let

$\ell_{n,m} := [\exp(m)/\alpha_n]$, $\ell_{n,M} := [\exp(M)/\alpha_n]$ and
$\ell_{n,x} := [\exp(x)/\alpha_n]$.

As a consequence, $\exp(x) \in [\ell_{n,x}\alpha_n - \alpha_n/2, \ell_{n,x}\alpha_n + \alpha_n/2)$. Denote two sets

$$A_n := \bigcup_{k=\ell_{n,m}}^{\ell_{n,M}} [k\alpha_n - \frac{\alpha_n}{2} + c_n\gamma_n, k\alpha_n + \frac{\alpha_n}{2} - c_n\gamma_n]$$

and

$$A_n^c := \bigcup_{k=\ell_{n,m}-1}^{\ell_{n,M}} (k\alpha_n + \frac{\alpha_n}{2} - c_n\gamma_n, k\alpha_n + \frac{\alpha_n}{2} + c_n\gamma_n).$$

Note that the Lebesgue measure of A_n^c goes to zero.

Lemma 1. *Under Assumption 1 and Condition RO(a)&(d), if $c_n \to \infty$ and $c_n/b_n \to 0$, then*

$$E\sum_{i=1}^{n_s} I_{A_n^c}(e^{X_{\tau_i}}) \le C \cdot n_s\frac{c_n}{b_n} \quad\text{and}\quad E\left(\sum_{i=1}^{n_s} I_{A_n^c}(e^{X_{\tau_i}})\right)^2 \le C\cdot\max\left\{n_s^2\frac{c_n^2}{b_n^2}, n_s\frac{c_n}{b_n}\right\},$$

where b_n is as defined in (26).

Proof. By the definition of A_n^c, we note that the Lebesgue measure $|A_n^c|$ of A_n^c is given by

$$|A_n^c| = 2(\ell_{n,M} - \ell_{n,m} + 2)\times c_n\gamma_n \le \frac{C}{\alpha_n}\times c_n\gamma_n = C\cdot\frac{c_n}{b_n}.$$

Notice also that

$$E\left(\sum_{i=1}^{n_s} I_{A_n^c}(e^{X_{\tau_i}})\right)^2 = E\sum_{i=1}^{n_s} I_{A_n^c}(e^{X_{\tau_i}}) + 2E\sum_{i=1}^{n_s}\sum_{j=i+1}^{n_s} I_{A_n^c}(e^{X_{\tau_i}})I_{A_n^c}(e^{X_{\tau_j}}).$$

Let $p_{S_{\tau_0}}(\cdot)$ be the marginal density of $S_{\tau_0} \equiv \exp(X_{\tau_0})$. By the strong Markov property of $(S_t)_{t\ge 0}$ and Condition RO(d) for the transition densities, we have that

$$\begin{aligned}
&E\sum_{i=1}^{n_s}\sum_{j=i+1}^{n_s} I_{A_n^c}(e^{X_{\tau_i}})I_{A_n^c}(e^{X_{\tau_j}})\\
&= \int\left(\sum_{i=1}^{n_s}\sum_{j=i+1}^{n_s}\int_{A_n^c}\left(\int_{A_n^c} p_{S_{\tau_j}|S_{\tau_i}}(z|y)dz\right)p_{S_{\tau_i}|S_{\tau_0}}(y|y_0)dy\right)\\
&\quad\times p_{S_{\tau_0}}(y_0)dy_0\\
&\le Cn_s(|A_n^c|)^2\sum_{i=1}^{n_s}\sum_{j=i+1}^{n_s}\frac{1}{\sqrt{i}}\frac{1}{\sqrt{j-i}}\\
&\le Cn_s(|A_n^c|)^2\int_0^{n_s}\frac{1}{\sqrt{x}}\left(\int_0^{n_s-x}\frac{1}{\sqrt{y}}dy\right)dx = C\cdot n_s^2\frac{c_n^2}{b_n^2}.
\end{aligned}$$

The same argument leads to $E\sum_{i=1}^{n_s} I_{A_n^c}(e^{X_{\tau_i}}) \le C\cdot n_s c_n/b_n$. Thus we are done. ■

Lemma 2. *Under Assumption 1 and Condition RO(a)&(d), if $c_n \to \infty$ and $c_n/b_n \to 0$, then*

$$\sup_{x\in[m,M]} E\left(R_{\tau_i}^2 I_{A_n}(e^{X_{\tau_i}})|X_{\tau_i} = x\right) \le 2\gamma_n^2 + \frac{C_1\alpha_n^2}{c_n}e^{-\frac{C_2c_n^2}{e^{2M}}}, \tag{27}$$

$$E\left(\sum_{i=1}^{n_s} R_{\tau_i}^2 I_{A_n^c}(e^{X_{\tau_i}})\right) \le C\alpha_n^2\times E\sum_{i=1}^{n_s} I_{A_n^c}(e^{X_{\tau_i}}) + 2n_s\gamma_n^2, \tag{28}$$

and consequently,

$$E\left(\sum_{i=1}^{n_s} R_{\tau_i}^2\right) \le Cn_s\gamma_n^2 + Cn_s\alpha_n^2\frac{c_n}{b_n} + \frac{C_1n_s\alpha_n^2}{c_n}e^{-\frac{C_2c_n^2}{e^{2M}}} = o(n_s\alpha_n^2). \tag{29}$$

Proof. We suppress the subscript of η by using η instead of η_{τ_i} for notational convenience. Recall that $\eta \sim N(0, \gamma_n^2)$ in Assumption 1. Let $\Phi(\cdot)$ be the standard Normal cumulative distribution function. Further define the following two intervals

$$\mathcal{I} := [\log\left(\frac{\ell_{n,x}\alpha_n - \alpha_n/2}{e^x}\right)/\gamma_n, \log\left(\frac{\ell_{n,x}\alpha_n + \alpha_n/2}{e^x}\right)/\gamma_n) \text{ and}$$

$$\widetilde{\mathcal{I}} = (-\infty, \log\left(\frac{3\alpha_n}{2}\right) - x),$$

and let $\mathcal{I}^c$ and $\widetilde{\mathcal{I}}^c$ be their complements, respectively. Note that

$$E(R_{\tau_i}^2) = E\left(R_{\tau_i}^2 I_{A_n}(e^{X_{\tau_i}})\right) + E\left(R_{\tau_i}^2 I_{A_n^c}(e^{X_{\tau_i}})\right). \tag{30}$$

(A):
In this part, we deal with $E(R_{\tau_i}^2 I_{A_n}(e^{X_{\tau_i}}))$ by establishing an upper bound for

$$\begin{aligned}
&\sup_{x\in[m,M]} E(R_{\tau_i}^2 I_{A_n}(e^{X_{\tau_i}})|X_{\tau_i} = x)\\
&= \sup_{x\in[m,M]} E(R_{\tau_i}^2 I_{A_n}(e^{X_{\tau_i}})I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}|X_{\tau_i} = x).
\end{aligned}\tag{31}$$

The above equality holds because given $X_{\tau_i} = x$, $R_{\tau_i} = \log\left((e^{x+\eta})^{(\alpha_n)}/(e^x)^{(\alpha_n)}\right) \equiv 0$ when $\eta/\gamma_n \in \mathcal{I}$. The right hand side of (31) can be bounded by

$$\begin{aligned}
&\sup_{x\in[m,M]} E(R_{\tau_i}^2 I_{A_n}(e^{X_{\tau_i}})I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}I_{\{\eta\in\widetilde{\mathcal{I}}\}}|X_{\tau_i} = x)\\
&+ \sup_{x\in[m,M]} E(R_{\tau_i}^2 I_{A_n}(e^{X_{\tau_i}})I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}I_{\{\eta\in\widetilde{\mathcal{I}}^c\}}|X_{\tau_i} = x).
\end{aligned}\tag{32}$$

The first term of (32) can be bounded by

$$\begin{aligned}
&\sup_{x\in[m,M]} E(R_{\tau_i}^2 I_{\{\eta\in\widetilde{\mathcal{I}}\}}|X_{\tau_i} = x)\\
&= \sup_{x\in[m,M]} \log\left(\frac{\alpha_n}{(e^x)^{(\alpha_n)}}\right)^2\Phi\left(\frac{\log\left(\frac{3\alpha_n}{2}\right) - x}{\gamma_n}\right)\\
&\le \log\left(\frac{\alpha_n}{(e^M)^{(\alpha_n)}}\right)^2\frac{\gamma_n}{\log\left(\frac{2e^m}{3\alpha_n}\right)}e^{-\frac{\log\left(\frac{2e^m}{3\alpha_n}\right)^2}{2\gamma_n^2}} = o(\alpha_n^2),
\end{aligned}\tag{33}$$

where the inequality follows by the tail property of standard Normal distribution $1 - \Phi(y) \le e^{-y^2/2}/y$ for $y > 0$. In fact, (33) goes to zero faster than α_n^r for any $r > 0$ and thus the first term in (32) can be completely ignored. *The same argument applies to the second term in (30) since the set A_n plays no role so far.*

Next, we consider the second term in (32). Denote

$$\overline{R_{\tau_i}} := \log\left(1 + \frac{e^x - (e^x)^{(\alpha_n)} + \frac{\alpha_n}{2e^\eta}}{(e^x)^{(\alpha_n)}}\right) + \eta \text{ and}$$

$$\underline{R_{\tau_i}} := \eta - \log\left(1 + \frac{\frac{\alpha_n}{2} + \frac{e^\eta\alpha_n}{2}}{(e^{x+\eta})^{(\alpha_n)}}\right).$$

When $\eta \in \widetilde{\mathcal{I}}^c$, by the definition of R_{τ_i}, we have (conditional on $X_{\tau_i} = x$) that

$$R_{\tau_i} \le \log\left(\frac{e^{\eta+x} + \alpha_n/2}{(e^x)^{(\alpha_n)}}\right) = \overline{R_{\tau_i}} \text{ and}$$

$$R_{\tau_i} \ge \log\left(\frac{(e^{x+\eta})^{(\alpha_n)}}{(e^{x+\eta})^{(\alpha_n)} + \frac{\alpha_n}{2} + \frac{e^\eta\alpha_n}{2}}\right) + \eta = \underline{R_{\tau_i}},$$

where the latter follows by noticing that $(e^{x+\eta})^{(\alpha_n)}+\alpha_n/2 \geq e^{x+\eta} \geq ((e^x)^{(\alpha_n)} - \alpha_n/2)e^{\eta}$. Thus from the last two inequalities, we can bound the second term in (32) by

$$\sup_{x\in[m,M]} E\overline{R_{\tau_i}}^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}} + \sup_{x\in[m,M]} E\underline{R_{\tau_i}}^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}I_{\{\eta\in\widetilde{\mathcal{I}}^c\}}. \tag{34}$$

Denote

$$\widetilde{\eta} := \log\left(1+\frac{e^x-(e^x)^{(\alpha_n)}+\frac{\alpha_n}{2e^{\eta}}}{(e^x)^{(\alpha_n)}}\right) \text{ and}$$

$$\breve{\eta} := -\log\left(1+\frac{\frac{\alpha_n}{2}+\frac{e^{\eta}\alpha_n}{2}}{(e^{x+\eta})^{(\alpha_n)}}\right).$$

Then the first term in (34) can be bounded by

$$\begin{aligned}&\sup_{x\in[m,M]} E\left(2\eta^2+2\widetilde{\eta}^2\right) I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}\\ &\quad\leq 2\gamma_n^2 + 2\sup_{x\in[m,M]} E\widetilde{\eta}^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}},\end{aligned} \tag{35}$$

where

$$\begin{aligned}&\sup_{x\in[m,M]} E\widetilde{\eta}^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}\\
\leq\ & \sup_{x\in[m,M]} 4\log(2)^2 E\left(\frac{e^x-(e^x)^{(\alpha_n)}+\frac{\alpha_n}{2e^{\eta}}}{(e^x)^{(\alpha_n)}}\right)^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}\\
=\ & \sup_{x\in[m,M]} 4\log(2)^2 I_{A_n}(e^x)\times\\
&\left(\left(\frac{e^x-(e^x)^{(\alpha_n)}}{(e^x)^{(\alpha_n)}}\right)^2 P\left(\frac{\eta}{\gamma_n}\in\mathcal{I}^c\right)+\left(\frac{\alpha_n}{2(e^x)^{(\alpha_n)}}\right)^2 Ee^{-2\eta}I_{\{\frac{\eta}{\gamma_n}\in\mathcal{I}^c\}}\right.\\
&\left.+\frac{\alpha_n(e^x-(e^x)^{(\alpha_n)})}{((e^x)^{(\alpha_n)})^2}Ee^{-\eta}I_{\{\frac{\eta}{\gamma_n}\in\mathcal{I}^c\}}\right)\\
\leq\ & \frac{\log(2)^2\alpha_n^2}{((e^m)^{(\alpha_n)})^2}\sup_{x\in[m,M]} P\left(\frac{\eta}{\gamma_n}\in\mathcal{I}^c\right)I_{A_n}(e^x)\\
&+\frac{\log(2)^2\alpha_n^2}{((e^m)^{(\alpha_n)})^2}\sup_{x\in[m,M]} Ee^{-2\eta}I_{\{\frac{\eta}{\gamma_n}\in\mathcal{I}^c\}}I_{A_n}(e^x)+\\
&\frac{2\log(2)^2\alpha_n^2}{((e^m)^{(\alpha_n)})^2}\sup_{x\in[m,M]} Ee^{-\eta}I_{\{\frac{\eta}{\gamma_n}\in\mathcal{I}^c\}}I_{A_n}(e^x).\end{aligned} \tag{36}$$

Notice that $P(\eta/\gamma_n \in \mathcal{I}^c) = E(e^{-0\times\eta}I_{\{\eta/\gamma_n\in\mathcal{I}^c\}})$. It suffices to evaluate for $r \geq 0$ the following

$$\begin{aligned}&\sup_{x\in[m,M]} Ee^{-r\eta}I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}I_{A_n}(e^x)\\
=\ & \sup_{x\in[m,M]}\int_{\mathcal{I}^c} e^{-r\gamma_n z}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz I_{A_n}(e^x)\\
\leq\ & \sup_{x\in[m,M]}\left(\int_{-\infty}^{\frac{-c_n}{e^M}} e^{-r\gamma_n z}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right.\\
&\left.+\int_{\frac{\log(2)c_n}{e^M}}^{\infty} e^{-r\gamma_n z}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right)I_{A_n}(e^x)\end{aligned} \tag{37}$$

$$\begin{aligned}\leq\ & e^{r^2\gamma_n^2/2}\left(\frac{e^M}{c_n-re^M\gamma_n}e^{-\frac{(c_n-re^M\gamma_n)^2}{2e^{2M}}}\right.\\
&\left.+\frac{e^M}{\log(2)c_n+re^M\gamma_n}e^{-\frac{(\log(2)c_n+re^M\gamma_n)^2}{2e^{2M}}}\right),\end{aligned} \tag{38}$$

where (37) follows from the fact that $I_{A_n}(e^x) = 1$ implies $\exp(x) \in [\ell_{n,x}\alpha_n - \alpha_n/2 + c_n\gamma_n, \ell_{n,x}\alpha_n + \alpha_n/2 - c_n\gamma_n]$ and hence

$$\left.\begin{aligned}\frac{\log\left(\frac{\ell_{n,x}\alpha_n-\alpha_n/2}{e^x}\right)}{\gamma_n} &\leq \frac{\ell_{n,x}\alpha_n-\alpha_n/2-e^x}{e^x\gamma_n}\\
&\leq \frac{-c_n\gamma_n}{e^M\gamma_n} = \frac{-c_n}{e^M}\to-\infty,\\
\frac{\log\left(\frac{\ell_{n,x}\alpha_n+\alpha_n/2}{e^x}\right)}{\gamma_n} &\geq \log(2)\frac{\ell_{n,x}\alpha_n+\alpha_n/2-e^x}{e^x\gamma_n}\\
&\geq \frac{\log(2)c_n\gamma_n}{e^M\gamma_n} = \frac{\log(2)c_n}{e^M}\to\infty;\end{aligned}\right\} \tag{39}$$

and (38) follows by the tail property of standard Normal distribution. From (35), (36) and (38), we have shown that

$$\sup_{x\in[m,M]} E\overline{R_{\tau_i}}^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}} \leq 2\gamma_n^2+\frac{C_1\alpha_n^2}{c_n}e^{-\frac{C_2c_n^2}{e^{2M}}}. \tag{40}$$

The second term in (34) can be bounded by

$$\begin{aligned}&\sup_{x\in[m,M]} E(2\eta^2+2\breve{\eta}^2)I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}I_{\{\eta\in\widetilde{\mathcal{I}}^c\}}\\
&\quad\leq 2\gamma_n^2+2\sup_{x\in[m,M]} E\breve{\eta}^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}I_{\{\eta\in\widetilde{\mathcal{I}}^c\}},\end{aligned}$$

where

$$\begin{aligned}&\sup_{x\in[m,M]} E\breve{\eta}^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}I_{\{\eta\in\widetilde{\mathcal{I}}^c\}}\\
&\leq \sup_{x\in[m,M]} E\left(\frac{\frac{\alpha_n}{2}+\frac{e^{\eta}\alpha_n}{2}}{(e^{m+\eta})^{(\alpha_n)}}\right)^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}\\
&\leq \frac{\alpha_n^2}{4}\sup_{x\in[m,M]} E\left(\frac{1+e^{\eta}}{(e^{m+\eta})^{(\alpha_n)}}\right)^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}\\
&\leq \alpha_n^2\sup_{x\in[m,M]} E\left(\frac{1+e^{\eta}}{e^{m+\eta}}\right)^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}.\end{aligned} \tag{41}$$

Hence, the same upper bound as in (40) holds for $\sup_{x\in[m,M]}E\underline{R_{\tau_i}}^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}I_{\{\eta\in\widetilde{\mathcal{I}}^c\}}$ by the same argument, i.e. (39) and the tail property of standard Normal distribution, used to get the upper bound for $\sup_{x\in[m,M]}E\overline{R_{\tau_i}}^2 I_{A_n}(e^x)I_{\{\eta/\gamma_n\in\mathcal{I}^c\}}$. This together with (31), (32), (33) and (34) leads to (27), i.e.,

$$\sup_{x\in[m,M]} E\left(R_{\tau_i}^2 I_{A_n}(e^{X_{\tau_i}})|X_{\tau_i}=x\right) \leq 2\gamma_n^2+\frac{C_1\alpha_n^2}{c_n}e^{-\frac{C_2c_n^2}{e^{2M}}}.$$

(B):
From (36) and (41), it is easy to see that

$$E\left(R_{\tau_i}^2 I_{A_n^c}(e^{X_{\tau_i}})|X_{\tau_i}\right) \leq 2\gamma_n^2 + C\alpha_n^2 I_{A_n^c}(e^{X_{\tau_i}}),$$

which leads to (28). Moreover, we have that

$$\begin{aligned}E\left(\sum_{i=1}^{n_s} R_{\tau_i}^2 I_{A_n^c}(e^{X_{\tau_i}})\right) &\leq C\alpha_n^2\times E\sum_{i=1}^{n_s} I_{A_n^c}(e^{X_{\tau_i}})+2n_s\gamma_n^2\\
&= Cn_s\alpha_n^2\frac{c_n}{b_n}+2n_s\gamma_n^2,\end{aligned}$$

where the last equality follows from Lemma 1. Hence, from (30), (27) and (28), we have that

$$E\left(\sum_{i=1}^{n_s} R_{\tau_i}^2\right) \leq Cn_s\gamma_n^2 + Cn_s\alpha_n^2\frac{c_n}{b_n}+\frac{C_1n_s\alpha_n^2}{c_n}e^{-\frac{C_2c_n^2}{e^{2M}}} = o(n_s\alpha_n^2),$$

because we let $c_n \to \infty$ and $c_n/b_n = o(1)$, completing the proof. ■

Proof of Proposition 1. We shall prove

$$[Y,Y]_1^{\mathcal{G}_s} - \int_0^1 \sigma_t^2 dt - \frac{\beta_\kappa^2}{6}\int_0^1 \frac{1}{S_t^2}dt + \frac{\beta_\kappa^2}{\pi^2}\int_0^1 \frac{1}{S_t^2}\sum_{k=1}^\infty \frac{1}{k^2}\exp\left(-2\pi^2k^2\frac{\sigma_t^2S_t^2}{\beta_\kappa^2}\right)dt \xrightarrow{p} 0.$$

The results of the Proposition then follows immediately by noticing that $n_s = n^{1-\kappa} + O(1)$.

Observe that

$$[Y,Y]_1^{\mathcal{G}_s} = [\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s} + [R,R]_1^{\mathcal{G}_s} + 2[\widetilde{Y},R]_1^{\mathcal{G}_s}. \tag{42}$$

By Li and Mykland (2015) we have that

$$[\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s} - \int_0^1 \sigma_t^2 dt - \frac{\beta_\kappa^2}{6}\int_0^1 \frac{1}{S_t^2}dt + \frac{\beta_\kappa^2}{\pi^2}\int_0^1 \frac{1}{S_t^2}\sum_{k=1}^\infty \frac{1}{k^2}\exp\left(-2\pi^2k^2\frac{\sigma_t^2S_t^2}{\beta_\kappa^2}\right)dt \xrightarrow{p} 0. \tag{43}$$

Therefore, we need to prove terms except $[\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s}$ in the right hand side of (42) are of order $o_p(1)$. We do this as follows. First,

$$[R,R]_1^{\mathcal{G}_s} = \sum_{i=0}^{n_s-1}(R_{\tau_{i+1}}^2 + R_{\tau_i}^2 - 2R_{\tau_i}R_{\tau_{i+1}}) \le 2\sum_{i=0}^{n_s-1}(R_{\tau_{i+1}}^2 + R_{\tau_i}^2) = 4\sum_{i=1}^{n_s}R_{\tau_i}^2 + 2R_{\tau_0}^2 - 2R_{\tau_{n_s}}^2.$$

Thus we have $[R,R]_1^{\mathcal{G}_s} = o_p(n_s\alpha_n^2)$ by Lemma 2 or specifically (29). It is easy to show $[\widetilde{Y},R]_1^{\mathcal{G}_s} = o_p(1)$ by applying (29) of Lemma 2 and the Cauchy–Schwarz inequality. However, for later use in Part B of the proof of Proposition 6, we derive the following (stronger) universal bound for the second moment of $[\widetilde{Y},R]_1^{\mathcal{G}_s}$. (see Eq. (44) in Box I)

By (27) and (43), and because we let $c_n \to \infty$,

$$(I) \le \left(Cn_s\gamma_n^2 + \frac{C_1 n_s\alpha_n^2}{c_n}e^{-\frac{C_2c_n^2}{e^{2M}}}\right)E[\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s} \le Cn_s\gamma_n^2 + \frac{C_1 n_s\alpha_n^2}{c_n}e^{-\frac{C_2c_n^2}{e^{2M}}} = o(1).$$

By (28), the Cauchy–Schwarz inequality and (43), and because we let $c_n \to \infty$ and $c_n/b_n = o(1)$,

$$\begin{aligned}(II) &\le C\alpha_n^2E\left([\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s}\sum_{i=1}^{n_s}I_{A_n^c}(e^{X_{\tau_i}})\right) + Cn_s\gamma_n^2E[\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s}\\
&\le C\alpha_n^2\left[E\left(\sum_{i=1}^{n_s}I_{A_n^c}(e^{X_{\tau_i}})\right)^2\right]^{1/2} + Cn_s\gamma_n^2\\
&\le C\alpha_n^2\max\left\{n_s\frac{c_n}{b_n}, n_s^{1/2}\frac{c_n^{1/2}}{b_n^{1/2}}\right\} + Cn_s\gamma_n^2 = o(1),\end{aligned}$$

where the third inequality follows from Lemma 1. Thus we are done. ■

Proof of Proposition 2. By Li and Mykland (2015) and $n_s = n^{1-\kappa} + O(1)$,

$$\sqrt{n^{1-\kappa}}\left([\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s} - \int_0^1 \sigma_t^2dt - \frac{n_s\alpha_n^2}{6}\int_0^1\frac{1}{S_t^2}dt\right) \to \int_0^1\sqrt{2}\sigma_t^2dB_t$$

stably in law, where B_t is a standard Brownian motion that is independent of the underlying σ-field. Then from (42), we need further to prove $\sqrt{n_s}([R,R]_1^{\mathcal{G}_s} + 2[\widetilde{Y},R]_1^{\mathcal{G}_s}) = o_p(1)$. It is easy to see from the proof of Proposition 1 that $2\sqrt{n_s}[\widetilde{Y},R]_1^{\mathcal{G}_s}$ dominates. From (44) (given in Box I) in the proof of Proposition 1 and the discussion immediately following it, we have that

$$\begin{aligned}E\left(\sqrt{n_s}[\widetilde{Y},R]_1^{\mathcal{G}_s}\right)^2 &\le Cn_s\left(n_s\gamma_n^2 + \alpha_n^2\max\left\{\frac{n_sc_n}{b_n}, \frac{n_s^{1/2}c_n^{1/2}}{b_n^{1/2}}\right\} + \frac{C_1n_s\alpha_n^2}{c_n}e^{-\frac{C_2c_n^2}{e^{2M}}}\right)\\
&= o(1)\end{aligned}$$

by letting $c_n = \log(b_n)$ and using the Condition RO(b). Thus we are done. ■

A.3. Proofs for the results in Section 2.3

We first introduce some new notations that will be used later. Let

$$\phi_{n,x}(z) := \frac{1}{\sqrt{2\pi}\gamma_n}e^{-(z-x)^2/(2\gamma_n^2)} \quad \text{and}$$

$$f_n(x) := \int_{-\infty}^{\infty}\log\left[\left(e^z\right)^{(\alpha_n)}\right]\phi_{n,x}(z)dz.$$

Note that $f_n(X_{\tau_i}) = E(Y_{\tau_i}|X_{\tau_i})$. We denote

$$\epsilon_{\tau_i} := Y_{\tau_i} - f_n(X_{\tau_i}),\ g_n(x) := E\left(\epsilon_{\tau_i}^2|X_{\tau_i} = x\right) \ \text{and}$$
$$G_n(x) := E\left(\epsilon_{\tau_i}^4|X_{\tau_i} = x\right).$$

Moreover, define

$$A(n,z) := \log\left((e^z)^{(\alpha_n)}\right) - z \text{ and } B(z,x) := z - x.$$

Lemma 3. *If Assumption 1 and Condition RA(a) are satisfied, then*

$$\sup_{x\in[m,M]}|f_n'(x) - 1| = O\left(\alpha_n/\gamma_n\right).$$

Proof. Recall the definitions $\ell_{n,x} := [\exp(x)/\alpha_n]$, $\ell_{n,m} := [\exp(m)/\alpha_n]$ and $\ell_{n,M} := [\exp(M)/\alpha_n]$. We have that $\ell_{n,m} \le \ell_{n,x} \le \ell_{n,M}$ and $1/\ell_{n,x} = O(\alpha_n)$ uniformly in x. For notational compactness, we denote the following consecutive intervals

$$\mathcal{I}_1 := \left(\frac{\log\left(\frac{\exp(x)}{3\alpha_n/2}\right)}{\gamma_n}, \infty\right) \text{ and}$$

$$\mathcal{I}_k := \left(\frac{\log\left(\frac{\exp(x)}{(k+1/2)\alpha_n}\right)}{\gamma_n}, \frac{\log\left(\frac{\exp(x)}{(k-1/2)\alpha_n}\right)}{\gamma_n}\right] \text{ for } k = 2,3,\ldots,\infty;$$

$$\widetilde{\mathcal{I}}_1 := (-\infty, \log(3\alpha_n/2)) \text{ and}$$
$$\widetilde{\mathcal{I}}_k := [\log(k\alpha_n - \alpha_n/2), \log(k\alpha_n + \alpha_n/2)) \text{ for } k = 2,3\ldots,\infty.$$

Then $\sup_{x\in[m,M]}|f_n'(x) - 1|$ can be bounded as follows

$$\begin{aligned}&\sup_{x\in[m,M]}|f_n'(x) - 1|\\
&= \sup_{x\in[m,M]}\left|\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\gamma_n}\left[\log\left((e^z)^{(\alpha_n)}\right) - z\right]e^{-(z-x)^2/(2\gamma_n^2)}\frac{z-x}{\gamma_n^2}dz\right|\\
&= \sup_{x\in[m,M]}\left|\sum_{k=1}^{\infty}\int_{\widetilde{\mathcal{I}}_k}\frac{1}{\sqrt{2\pi}\gamma_n}\left[\log(k\alpha_n) - z\right]e^{-(z-x)^2/(2\gamma_n^2)}\frac{z-x}{\gamma_n^2}dz\right|\\
&= \sup_{x\in[m,M]}\left|\sum_{k=1}^{\infty}\int_{\mathcal{I}_k}\frac{1}{\sqrt{2\pi}\gamma_n}\left(\log\left(\frac{\exp(x)}{k\alpha_n}\right) - \gamma_nz\right)e^{-z^2/2}zdz\right|\\
&\le \sup_{x\in[m,M]}\frac{2\log(2)\alpha_n}{\exp(x)\gamma_n}\left(\sum_{k=1}^{\ell_{n,x}-1}\int_{\mathcal{I}_k}\frac{1}{\sqrt{2\pi}}e^{2\gamma_nz}e^{-z^2/2}|z|dz\right.\end{aligned}$$

$$
\begin{aligned}
&E\left([\widetilde{Y},R]_1^{\mathcal{G}_s}\right)^2 \leq E\left(\sum_{i=0}^{n_s-1}(\Delta\widetilde{Y}_{\tau_i})^2\sum_{i=0}^{n_s-1}(\Delta R_{\tau_i})^2\right) \leq E\left([\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s}E\left[4\sum_{i=1}^{n_s}R_{\tau_i}^2+2R_{\tau_0}^2-2R_{\tau_{n_s}}^2|X\right]\right)\\
&=\underbrace{E\left([\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s}E\left[4\sum_{i=1}^{n_s}R_{\tau_i}^2I_{A_n}(e^{X_{\tau_i}})+2R_{\tau_0}^2I_{A_n}(e^{X_{\tau_0}})-2R_{\tau_{n_s}}^2I_{A_n}(e^{X_{\tau_{n_s}}})|X\right]\right)}_{(I)}\\
&\quad+\underbrace{E\left([\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_s}E\left[4\sum_{i=1}^{n_s}R_{\tau_i}^2I_{A_n^c}(e^{X_{\tau_i}})+2R_{\tau_0}^2I_{A_n^c}(e^{X_{\tau_0}})-2R_{\tau_{n_s}}^2I_{A_n^c}(e^{X_{\tau_{n_s}}})|X\right]\right)}_{(II)} \qquad (44)
\end{aligned}
$$

Box I.

$$
\begin{aligned}
&+\sum_{\ell_{n,x}}^{\infty}\int_{\mathcal{I}_k}\frac{1}{\sqrt{2\pi}}e^{\gamma_n z}e^{-z^2/2}|z|dz\Bigg) \qquad (45)\\
&\leq \frac{2\log(2)\alpha_n}{\exp(m)\gamma_n}\left(\int_{-\infty}^{\infty}e^{2\gamma_n z}|z|\phi(z)dz+\int_{-\infty}^{\infty}e^{\gamma_n z}|z|\phi(z)dz\right)\\
&=O\left(\frac{\alpha_n}{\gamma_n}\right),
\end{aligned}
$$

where in (45) we use the fact that

$$
\begin{cases}
\dfrac{\exp(x)}{\alpha_n}\left|\log(\dfrac{\exp(x)}{k\alpha_n})-\gamma_n z\right| \leq \dfrac{\exp(x)\log(2)}{k\alpha_n} \leq 2\log(2)e^{\gamma_n z},\\
\quad \text{for } z\in\mathcal{I}_k,\ k=\ell_{n,x},\dots,\infty;\\
\dfrac{\exp(x)}{\alpha_n}|\log\left(\dfrac{\exp(x)}{\alpha_n}\right)-\gamma_n z| \leq \dfrac{\exp(x)}{\alpha_n}\gamma_n z \leq 2\log(2)e^{2\gamma_n z},\\
\quad \text{for } z\in(\dfrac{\log\left(\frac{\exp(x)}{\frac{\alpha_n}{2}}\right)}{\gamma_n},\infty); \qquad (46)\\
\dfrac{\exp(x)}{\alpha_n}\left|\log(\dfrac{\exp(x)}{k\alpha_n})-\gamma_n z\right| \leq \dfrac{\exp(x)\log(2)}{k\alpha_n} \leq 2\log(2)e^{2\gamma_n z},\\
\quad \text{for } z\in\mathcal{I}_1\setminus(\dfrac{\log\left(\frac{\exp(x)}{\frac{\alpha_n}{2}}\right)}{\gamma_n},\infty)\ \&\ k=1,\\
\qquad\qquad \text{and } z\in\mathcal{I}_k,\ k=2,\dots,\ell_{n,x}-1.
\end{cases}
$$

$\phi(z)$ is the standard normal density. Thus we are done. ■

Lemma 4. *If Assumption 1 and Condition RA(a) are satisfied, then*

$$\sup_{x\in[m,M]}|f_n(x)-x|=O(\alpha_n).$$

Proof. We adopt the same notations $\ell_{n,x}$, $\ell_{n,m}$, $\ell_{n,M}$, $\mathcal{I}_k$, $k=1,2,3,\dots,\infty$, and $\widetilde{\mathcal{I}}_k$, $k=1,2,3\dots,\infty$, as in the proof for Lemma 3. Then $\sup_{x\in[m,M]}|f_n(x)-x|$ can be bounded as follows

$$
\begin{aligned}
&\sup_{x\in[m,M]}|f_n(x)-x|\\
&=\sup_{x\in[m,M]}\left|\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\gamma_n}\left[\log\left((e^z)^{(\alpha_n)}\right)-z\right]e^{-(z-x)^2/(2\gamma_n^2)}dz\right|\\
&=\sup_{x\in[m,M]}\left|\sum_{k=1}^{\infty}\int_{\widetilde{\mathcal{I}}_k}\frac{1}{\sqrt{2\pi}\gamma_n}(z-\log(k\alpha_n))e^{-(z-x)^2/(2\gamma_n^2)}dz\right|\\
&=\sup_{x\in[m,M]}\left|\sum_{k=1}^{\infty}\int_{\mathcal{I}_k}\frac{1}{\sqrt{2\pi}}\left(\log\left(\frac{\exp(x)}{k\alpha_n}\right)-\gamma_n z\right)e^{-z^2/2}dz\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sup_{x\in[m,M]}\frac{2\log(2)\alpha_n}{\exp(x)}\left(\sum_{k=1}^{\ell_{n,x}-1}\int_{\mathcal{I}_k}\frac{1}{\sqrt{2\pi}}e^{2\gamma_n z}e^{-z^2/2}dz\right.\\
&\quad\left.+\sum_{k=\ell_{n,x}}^{\infty}\int_{\mathcal{I}_k}\frac{1}{\sqrt{2\pi}}e^{\gamma_n z}e^{-z^2/2}dz\right) \qquad (47)\\
&\leq \frac{2\log(2)\alpha_n}{\exp(m)}\left(\int_{-\infty}^{\infty}e^{2\gamma_n z}\phi(z)dz+\int_{-\infty}^{\infty}e^{\gamma_n z}\phi(z)dz\right)=O(\alpha_n),
\end{aligned}
$$

where in (47) we use the same inequalities as in (46). Thus we are done. ■

Lemma 5. *If Assumption 1 and Condition RA(a) are satisfied, then*

$$\sup_{x\in[m,M]}|\int A(n,z)^i\phi_{n,x}(z)dz|=O(\alpha_n^i)$$

for $i\geq 1$.

Proof. This is a straightforward generalization of Lemma 4. ■

Lemma 6. *If Assumption 1 and Condition RA(a) are satisfied, then*

$$\sup_{x\in[m,M]}|\int A(n,z)^iB(z,x)^j\phi_{n,x}(z)dz|=O(\alpha_n^i\gamma_n^j)$$

for $i\geq 1$ and $j\geq 1$.

Proof. This is a straightforward generalization of Lemma 3. ■

Proof of Proposition 3. Observing that

$$[Y,Y]_1^{\mathcal{G}_s}=[\epsilon,\epsilon]_1^{\mathcal{G}_s}+[f_n(X),f_n(X)]_1^{\mathcal{G}_s}+2[f_n(X),\epsilon]_1^{\mathcal{G}_s}, \qquad (48)$$

we divide the proof into the following three parts.
Part (a):
We show the negligibility in probability of the following term

$$
\begin{aligned}
[f_n(X),f_n(X)]_1^{\mathcal{G}_s}-\int_0^1\sigma_t^2dt&=\sum_{i=1}^{n_s}(f_n'(\zeta_i)^2-1)(X_{\tau_i}-X_{\tau_{i-1}})^2\\
&\quad+[X,X]_1^{\mathcal{G}_s}-\int_0^1\sigma_t^2dt,
\end{aligned}
$$

where by the mean value theorem ζ_i lies between $X_{\tau_{i-1}}$ and X_{τ_i}. By Lemma 3

$$\left|\sum_{i=1}^{n_s}(f_n'(\zeta_i)^2-1)(X_{\tau_i}-X_{\tau_{i-1}})^2\right|$$
$$\le \left(\sup_{x\in[m,M]}|f_n'(x)-1|+2\right)\sup_{x\in[m,M]}|f_n'(x)-1|\times[X,X]_1^{\mathcal{G}_s}$$
$$\le C\sup_{x\in[m,M]}|f_n'(x)-1|\times[X,X]_1^{\mathcal{G}_s}=O_p\left(\frac{\alpha_n}{\gamma_n}\right)$$

as $[X,X]_1^{\mathcal{G}_s}=O_p(1)$ because $[X,X]_1^{\mathcal{G}_s}-\int_0^1\sigma_t^2dt\xrightarrow{p}0$. Hence, we proved $[f_n(X),f_n(X)]_1^{\mathcal{G}_s}-\int_0^1\sigma_t^2dt\xrightarrow{p}0$ under the assumption $\alpha_n/\gamma_n\to 0$.
Part (b):
We prove convergence to zero in probability of the following term

$$[\epsilon,\epsilon]_1^{\mathcal{G}_s}-2n_s\gamma_n^2=2\sum_{i=1}^{n_s}(\epsilon_{\tau_i}^2-E(\epsilon_{\tau_i}^2|X))-2\sum_{i=1}^{n_s}\epsilon_{\tau_i}\epsilon_{\tau_{i-1}}-\epsilon_{\tau_{n_s}}^2$$
$$+\epsilon_{\tau_0}^2+2\sum_{i=1}^{n_s}E(\epsilon_{\tau_i}^2|X)-2n_s\gamma_n^2. \tag{49}$$

Observe that

$$2\sum_{i=1}^{n_s}E(\epsilon_{\tau_i}^2|X)-2n_s\gamma_n^2=2\sum_{i=1}^{n_s}g_n(X_{\tau_i})-2n_s\gamma_n^2$$
$$=-2\sum_{i=1}^{n_s}(f_n(X_{\tau_i})-X_{\tau_i})^2+2\sum_{i=1}^{n_s}\int_{-\infty}^{\infty}A(n,z)^2\phi_{n,X_{\tau_i}}(z)dz$$
$$+4\sum_{i=1}^{n_s}\int_{-\infty}^{\infty}A(n,z)B(z,x)\phi_{n,X_{\tau_i}}(z)dz$$

because $g_n(x)$ is given by

$$\int_{-\infty}^{\infty}\left[\log\left((e^z)^{(\alpha_n)}\right)-f_n(x)\right]^2\phi_{n,x}(z)dz$$
$$=\int_{-\infty}^{\infty}[A(n,z)+B(z,x)-(f_n(x)-x)]^2\phi_{n,x}(z)dz$$
$$=\gamma_n^2-(f_n(x)-x)^2+\int_{-\infty}^{\infty}A(n,z)^2\phi_{n,x}(z)dz$$
$$+2\int_{-\infty}^{\infty}A(n,z)B(z,x)\phi_{n,x}(z)dz. \tag{50}$$

By Lemma 4, $\sup_{x\in[m,M]}(f_n(x)-x)^2=O(\alpha_n^2)$. Notice that $\int_{-\infty}^{\infty}A(n,z)B(z,x)\phi_{n,x}(z)dz=\gamma_n^2(f_n'(x)-1)$. Hence, we have $\sup_{x\in[m,M]}|\int_{-\infty}^{\infty}A(n,z)B(z,x)\phi_{n,x}(z)dz|=O(\gamma_n\alpha_n)$. Moreover, by Lemma 5, $\sup_{x\in[m,M]}|\int_{-\infty}^{\infty}A(n,z)^2\phi_{n,x}(z)dz|=O(\alpha_n^2)$. Thus we proved $2\sum_{i=1}^{n_s}E(\epsilon_{\tau_i}^2|X)-2n_s\gamma_n^2=o_p(1)$ under the assumption of Proposition 3.

$-\epsilon_{\tau_{n_s}}^2+\epsilon_{\tau_0}^2=o_p(1)$ since $\sup_{x\in[m,M]}|g_n(x)|=O(\gamma_n^2)=o(1)$ from the calculation for $g_n(x)$.

$2\sum_{i=1}^{n_s}(\epsilon_{\tau_i}^2-E(\epsilon_{\tau_i}^2|X))=o_p(1)$ and $-2\sum_{i=1}^{n_s}\epsilon_{\tau_i}\epsilon_{\tau_{i-1}}=o_p(1)$ by considering their variances and $n_s\sup_{x\in[m,M]}|g_n(x)^2|=O(n_s\gamma_n^4)=o(1)$ from the calculation for $g_n(x)$ and $n_s\sup_{x\in[m,M]}|G_n(x)|=O(n_s\gamma_n^4)=o(1)$ which follows from Lemmas 5 & 6 and a similar calculation for $G_n(x)$ to that for $g_n(x)$. We refer to the proof of Lemma 7 for the detailed calculation of $G_n(x)$.

Thus we have proved $[\epsilon,\epsilon]_1^{\mathcal{G}_s}-2n_s\gamma_n^2\xrightarrow{p}0$.

Part (c):

$$E\left\{\left([f_n(X),\epsilon]_1^{\mathcal{G}_s}\right)^2|X\right\}$$
$$=\sum_{i=1}^{n_s-1}\left(\Delta f_n(X_{\tau_{i-1}})-\Delta f_n(X_{\tau_i})\right)^2E(\epsilon_{\tau_i}^2|X)$$
$$+\Delta f_n(X_{\tau_{n_s-1}})^2E(\epsilon_{\tau_{n_s}}^2|X)+\Delta f_n(X_{\tau_0})^2E(\epsilon_{\tau_0}^2|X)$$
$$\le\sup_{x\in[m,M]}g_n(x)\left[\sum_{i=1}^{n_s-1}\left(\Delta f_n(X_{\tau_{i-1}})-\Delta f_n(X_{\tau_i})\right)^2\right.$$
$$\left.+\Delta f_n(X_{\tau_{n_s-1}})^2+\Delta f_n(X_{\tau_0})^2\right]$$
$$=2\sup_{x\in[m,M]}g_n(x)\left([f_n(X),f_n(X)]_1^{\mathcal{G}_s}-\sum_{i=1}^{n_s-1}\Delta f_n(X_{\tau_{i-1}})\Delta f_n(X_{\tau_i})\right)$$
$$\le 4\sup_{x\in[m,M]}g_n(x)[f_n(X),f_n(X)]_1^{\mathcal{G}_s}=o_p(1)$$

by Part (a) and the calculation for $g_n(x)$ in Part (b), completing the proof. ■

Before we prove Proposition 4, we need the following Lemma 7. Let

$$M_1^{(1)}:=2\sqrt{n_s}\sum_{i=1}^{n_s}(\epsilon_{\tau_i}^2-E(\epsilon_{\tau_i}^2|X)),\ M_1^{(2)}:=2\sqrt{n_s}\sum_{i=1}^{n_s}\epsilon_{\tau_i}\epsilon_{\tau_{i-1}},$$

and

$$M_1^{(3)}:=2\sqrt{n_s}\sum_{i=1}^{n_s-1}\left(\Delta f_n(X_{\tau_{i-1}})-\Delta f_n(X_{\tau_i})\right)\epsilon_{\tau_i}.$$

Lemma 7. *Under the conditions of Proposition 4, we have that* $\langle M^{(1)},M^{(1)}\rangle_1\xrightarrow{p}8\gamma_\kappa^4$, $\langle M^{(2)},M^{(2)}\rangle_1\xrightarrow{p}4\gamma_\kappa^4$, and $\langle M^{(3)},M^{(3)}\rangle_1\xrightarrow{p}8\gamma_\kappa^2\int_0^1\sigma_t^2dt$.

Proof. First,

$$\langle M^{(1)},M^{(1)}\rangle_1=4n_s\sum_{i=1}^{n_s}\left(E(\epsilon_{\tau_i}^4|X)-(E(\epsilon_{\tau_i}^2|X))^2\right)$$
$$=4n_s\sum_{i=1}^{n_s}(G_n(X_{\tau_i})-g_n(X_{\tau_i})^2)$$
$$=8n_s^2\gamma_n^4+o_p(n_s^2\gamma_n^4)\xrightarrow{p}8\gamma_\kappa^4,$$

because $\sup_{x\in[m,M]}g_n(x)^2=\gamma_n^4+o(\gamma_n^4)$ by Eq. (50) and the discussion immediately following it; and

$$G_n(x)=\int_{-\infty}^{\infty}[A(n,z)+B(z,x)-(f_n(x)-x)]^4\phi_{n,x}(z)dz$$
$$=\int_{-\infty}^{\infty}B(z,x)^4\phi_{n,x}(z)dz+\int_{-\infty}^{\infty}A(n,z)^4\phi_{n,x}(z)dz-3(f_n(x)-x)^4$$
$$+6\int_{-\infty}^{\infty}\left[A(n,z)^2B(z,x)^2\right.$$
$$\left.+(f_n(x)-x)^2\left(A(n,z)^2+B(z,x)^2\right)\right]\phi_{n,x}(z)dz$$
$$+4\int_{-\infty}^{\infty}\left[A(n,z)^3B(z,x)+A(n,z)B(z,x)^3\right.$$
$$\left.-A(n,z)^3(f_n(x)-x)\right]\phi_{n,x}(z)dz$$
$$+12\int\left[(f_n(x)-x)^2A(n,z)B(z,x)-(f_n(x)-x)\right.$$

$$\times\ \left(A(n,z)^2B(z,x)+A(n,z)B(z,x)^2\right)\Big]\,\phi_{n,x}(z)dz$$
$$=3\gamma_n^4+\text{remainder terms},$$

where the "remainder terms" are of order $o_p(\gamma_n^4)$ by Lemmas 4–6.

Second,

$$\langle M^{(2)},M^{(2)}\rangle_1=4n_s\sum_{i=1}^{n_s}\left(\epsilon^2_{\tau_{i-1}}-g_n(X_{\tau_{i-1}})\right)g_n(X_{\tau_i})$$
$$+4n_s\sum_{i=1}^{n_s}g_n(X_{\tau_{i-1}})g_n(X_{\tau_i})$$
$$=4n_s^2\gamma_n^4+o_p(n_s^2\gamma_n^4)\xrightarrow{p}4\gamma_\kappa^4,$$

because $4n_s\sum_{i=1}^{n_s}g_n(X_{\tau_{i-1}})g_n(X_{\tau_i})=4n_s^2\gamma_n^4+o_p(n_s^2\gamma_n^4)$ by Eq. (50) and the discussion immediately following it; and

$$E\left(\left(4n_s\sum_{i=1}^{n_s}\left(\epsilon^2_{\tau_{i-1}}-g_n(X_{\tau_{i-1}})\right)g_n(X_{\tau_i})\right)^2|X\right)$$
$$=16n_s^2\sum_{i=1}^{n_s}\left(G_n(X_{\tau_{i-1}})-g_n(X_{\tau_{i-1}})^2\right)g_n(X_{\tau_i})^2$$
$$=32n_s^3\gamma_n^8+o_p(n_s^3\gamma_n^8)=o_p(1)$$

by the same calculation for $G_n(x)$ as above and $g_n(x)$ in Eq. (50).

Third, we compute the quadratic variation (conditional on X) of $M^{(3)}$ as follows

$$\langle M^{(3)},M^{(3)}\rangle_1=4n_s\sum_{i=1}^{n_s-1}\left(\Delta f_n(X_{\tau_{i-1}})-\Delta f_n(X_{\tau_i})\right)^2g_n(X_{\tau_i})$$
$$=8n_s\gamma_n^2\sum_{i=0}^{n_s-1}(\Delta f_n(X_{\tau_i}))^2-8n_s\gamma_n^2\sum_{i=1}^{n_s-1}\Delta f_n(X_{\tau_{i-1}})\Delta f_n(X_{\tau_i})$$
$$-4n_s\gamma_n^2(\Delta f_n(X_{\tau_0}))^2$$
$$-4n_s\gamma_n^2(\Delta f_n(X_{\tau_{n_s-1}}))^2+o_p(n_s\gamma_n^2),$$

where we use $\sup_{x\in[m,M]}|g_n(x)|=\gamma_n^2+o(\gamma_n^2)$ by Eq. (50). Notice that $-4n_s\gamma_n^2(\Delta f_n(X_{\tau_0}))^2=O_p(\gamma_n^2)=o_p(1)$ and $4n_s\gamma_n^2(\Delta f_n(X_{\tau_{n_s-1}}))^2=O_p(\gamma_n^2)=o_p(1)$. The expectation of the quadratic variation of the end point of martingale $-8n_s\gamma_n^2\sum_{i=1}^{n_s-1}\Delta f_n(X_{\tau_{i-1}})\Delta f_n(X_{\tau_i})$ is:

$$E\left(64n_s^2\gamma_n^4\sum_{i=1}^{n_s-1}(\Delta f_n(X_{\tau_{i-1}}))^2\int_{\tau_i}^{\tau_{i+1}}f_n'(X_t)^2\sigma_t^2dt\right)\le C64n_s\gamma_n^4$$

by Lemma 3 and the boundedness of σ_t. Hence, by the Burkholder–Davis–Gundy inequality $-8n_s\gamma_n^2\sum_{i=1}^{n_s-1}\Delta f_n(X_{\tau_{i-1}})\Delta f_n(X_{\tau_i})=O_p(n_s^{1/2}\gamma_n^2)=o_p(1)$. Lastly,

$$8n_s\gamma_n^2\sum_{i=0}^{n_s-1}(\Delta f_n(X_{\tau_i}))^2=8n_s\gamma_n^2[f_n(X),f_n(X)]_1^{\mathcal{G}_s}\xrightarrow{p}8\gamma_\kappa^2\int_0^1\sigma_t^2dt,$$

by Part (a) of the proof of Proposition 3. Thus we have

$$\langle M^{(3)},M^{(3)}\rangle_1=8n_s\gamma_n^2\sum_{i=0}^{n_s-1}(\Delta f_n(X_{\tau_i}))^2+o_p(1)\xrightarrow{p}8\gamma_\kappa^2\int_0^1\sigma_t^2dt.\quad\blacksquare$$

Now we are ready to present the proof of Proposition 4.

Proof of Proposition 4. The proof of Proposition 4 is based on the same decomposition (48) in the proof of Proposition 3. Accordingly, we also divide the proof here into the following four parts. Because $n_s=n^{1-\kappa}+O(1)$, it suffices to prove the limit result of Proposition 4 with $(n^{1-\kappa})^{1/2}$ being replaced by $n_s^{1/2}$.

(A):

From Part (a) of the proof of Proposition 3, the end point of martingale due to X

$$M_1:=\sqrt{n_s}([X,X]_1^{\mathcal{G}_s}-\int_0^1\sigma_t^2dt)\to\int_0^1(2\sigma_t^4)^{1/2}dB_t\tag{51}$$

stably in law by e.g. Mykland and Zhang (2006), and the remainder term $\sqrt{n_s}\sum_{\tau_i\in\mathcal{G}_s}(f_n'(\zeta_i)^2-1)(X_{\tau_i}-X_{\tau_{i-1}})^2$ is of order $O_p(\sqrt{n_s}\alpha_n/\gamma_n)=o_p(1)$ by the assumption of Proposition 4.

(B):

Next, we deal with

$$\sqrt{n_s}\left([\epsilon,\epsilon]_1^{\mathcal{G}_s}-2n_s\gamma_n^2\right).$$

We follow the decomposition (49) in Part (b) of the proof of Proposition 3, $-\epsilon^2_{\tau_{n_s}}+\epsilon^2_{\tau_0}=O_p(\sqrt{n_s}\gamma_n^2)=o_p(1)$ and $\sqrt{n_s}(2\sum_{\tau_i\in\mathcal{G}_s}E(\epsilon^2_{\tau_i}|X)-2n_s\gamma_n^2)=O_p(n_s^{3/2}\alpha_n\gamma_n)=o_p(1)$ by the calculation therein. This implies that the two end points of martingales due to random noise are the quantities dominate in this Part. They are

$$M_1^{(1)}=2\sqrt{n_s}\sum_{i=1}^{n_s}(\epsilon^2_{\tau_i}-E(\epsilon^2_{\tau_i}|X))\ \text{and}\ M_1^{(2)}=2\sqrt{n_s}\sum_{i=1}^{n_s}\epsilon_{\tau_i}\epsilon_{\tau_{i-1}}.$$

We compute their quadratic variations (conditional on X). By Lemma 7,

$$\langle M^{(1)},M^{(1)}\rangle_1\xrightarrow{p}8\gamma_\kappa^4\ \text{and}\ \langle M^{(2)},M^{(2)}\rangle_1\xrightarrow{p}4\gamma_\kappa^4.$$

(C):

From Part (c) of the proof of Proposition 3, we know that $2[f_n(X),\epsilon]_1^{\mathcal{G}_s}=O_p(\gamma_n)$. We need to investigate this term closely as follows

$$2\sqrt{n_s}[f_n(X),\epsilon]_1^{\mathcal{G}_s}=2\sqrt{n_s}\sum_{i=1}^{n_s-1}\left(\Delta f_n(X_{\tau_{i-1}})-\Delta f_n(X_{\tau_i})\right)\epsilon_{\tau_i}$$
$$-2\sqrt{n_s}\Delta f_n(X_{\tau_0})\epsilon_{\tau_0}+2\sqrt{n_s}\Delta f_n(X_{\tau_{n_s-1}})\epsilon_{\tau_{n_s}}.$$

Since $\Delta f_n(X_{\tau_i})=f_n'(\zeta_i)\Delta X_{\tau_i}=O_p(1/n_s^{1/2})$ by Lemma 3 where ζ_i lies between X_{τ_i} and $X_{\tau_{i+1}}$, we can easily see that $-2\sqrt{n_s}\Delta f_n(X_{\tau_0})\epsilon_{\tau_0}=O_p(\gamma_n)=o_p(1)$ and $2\sqrt{n_s}\Delta f_n(X_{\tau_{n_s-1}})\epsilon_{\tau_{n_s}}=O_p(\gamma_n)=o_p(1)$. It thus reduces to studying the limiting behavior of the following end point of martingale

$$M_1^{(3)}=2\sqrt{n_s}\sum_{i=1}^{n_s-1}\left(\Delta f_n(X_{\tau_{i-1}})-\Delta f_n(X_{\tau_i})\right)\epsilon_{\tau_i}.$$

By Lemma 7, the quadratic variation (conditional on X) of $M^{(3)}$ is

$$\langle M^{(3)},M^{(3)}\rangle_1\xrightarrow{p}8\gamma_\kappa^2\int_0^1\sigma_t^2dt.$$

(D):

Furthermore, we need to investigate the quadratic covariations (conditional on X) among $M^{(1)}$, $M^{(2)}$, $M^{(3)}$ and M. Obviously, $\langle M^{(1)},M^{(2)}\rangle_1=0$, $\langle M^{(1)},M^{(3)}\rangle_1=0$, $\langle M^{(1)},M\rangle_1=0$, $\langle M^{(2)},M\rangle_1=0$ and $\langle M^{(3)},M\rangle_1=0$. The quadratic covariation (conditional on X) $\langle M^{(2)},M^{(3)}\rangle_1$ is

$$\langle M^{(2)},M^{(3)}\rangle_1=4n_s\sum_{i=1}^{n_s-1}(\Delta f_n(X_{\tau_{i-1}})-\Delta f_n(X_{\tau_i}))g_n(X_{\tau_i})\epsilon_{\tau_{i-1}}.$$

The conditional (on X) second moment of $\langle M^{(2)}, M^{(3)}\rangle_1$ is

$$
\begin{aligned}
&E\left(\left(\langle M^{(2)}, M^{(3)}\rangle_1\right)^2 |X\right)\\
&= 16n_s^2 \sum_{i=1}^{n_s-1}(\Delta f_n(X_{\tau_{i-1}}) - \Delta f_n(X_{\tau_i}))^2 g_n(X_{\tau_i})^2 g_n(X_{\tau_{i-1}})\\
&= 16n_s^2 \sum_{i=1}^{n_s-1}(\Delta f_n(X_{\tau_{i-1}}) - \Delta f_n(X_{\tau_i}))^2(\gamma_n^6 + o_p(\gamma_n^6)) = O_p(n_s^2\gamma_n^6),
\end{aligned}
$$

where we use $\sup_{x\in[m,M]}|g_n(x)| = \gamma_n^2 + o(\gamma_n^2)$ by (50) in the second equality. The last equality follows from the fact that $16n_s^2\sum_{i=1}^{n_s}(\Delta f_n(X_{\tau_{i-1}}) - \Delta f_n(X_{\tau_i}))^2\gamma_n^6 \le 64n_s^2\gamma_n^6[f_n(X), f_n(X)]_1^{\mathcal{G}_s}$. Hence, $\langle M^{(2)}, M^{(3)}\rangle_1 \xrightarrow{p} 0$. The conditional Lyapunov conditions are satisfied by using the same yet tedious calculation as that for the previous quadratic (co)variations. Then (51), Lemma 7 and the martingale central limit theorem conclude the convergence result of Proposition 4 in view of, e.g., Hall and Heyde (1998) (Corollary 3.1, p.58) and Mykland and Zhang (2012). ■

A.4. Proof of Proposition 5

This is a direct consequence of Propositions 2 and 4.

A.5. Proofs for the results in Section 3

Before we prove Proposition 6, we need an additional lemma that is given below.

Lemma 8. *If Condition RO(c) holds and $\sqrt{n}\alpha_n = O(n^{-\varsigma})$ for some $\varsigma > 0$, then*

$$
E\sqrt{n}\left|[\widetilde{Y}, \widetilde{Y}]_1^{(\mathrm{all})} - [X, X]_1^{(\mathrm{all})} - \frac{n\alpha_n^2}{6}\int_0^1 \frac{1}{S_t^2}dt\right| \le C\cdot o(1),
$$

where C is a universal constant independent of the grid of times used.

Proof. In the following proof, rounding off is used. By the same argument of Li and Mykland (2015), the results also apply to the case of rounding to the nearest integer.

By the argument from equation (A.6) to (A.8) in Subsection A.3 of Li and Mykland (2015) and apply Delattre and Jacod (1997),

$$
E\sqrt{n}\left|[\widetilde{Y}, \widetilde{Y}]_1 - \sum_{i=1}^{n}\left(\frac{\widetilde{S}_{t_i} - \widetilde{S}_{t_{i-1}}}{\widetilde{S}_{t_{i-1}}}\right)^2\right| \le C\cdot o(1),
$$

and

$$
E\sqrt{n}\left|[X, X]_1 - \sum_{i=1}^{n}\left(\frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}}\right)^2\right| \le C\cdot o(1).
$$

Moreover, by applying the Burkholder–Davis–Gundy inequality, we have

$$
E\sqrt{n}\left|\frac{\beta_n^2}{6}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{S_{t_{i-1}}^2} - \frac{\beta_n^2}{6}\int_0^1\frac{1}{S_t^2}dt\right| \le C\cdot\beta_n^2,
$$

by $\exp(m) \le S_t \le \exp(M)$ for $t \in [0, 1]$ and boundedness of σ_t, where $\beta_n = \sqrt{n}\alpha_n$.

By the argument of Lemma 2 in Li and Mykland (2015), for any m and M, there exist $N_{m,M}$ large and $c_{m,M}$ such that for all $n \ge N_{m,M}$ and $i = 0, 1, \ldots, n$, $S_{t_i} \ge c_{m,M}$, $c_{m,M} \le \exp(m)$.

To be consistent with the notation of Delattre and Jacod (1997), with a little abuse of notation, we define a sequence of *tri-variate* functions

$$
\begin{aligned}
f_n(x, u, y) &:= \varphi_n(x - \alpha_n u, \beta_n\lfloor u + y/\beta_n\rfloor) - \varphi_n(x, y)\\
&\quad - \frac{\beta_n^2}{6}\varphi_n(x, 1), \text{ for } (x, u, v) \in \mathbb{R}^+ \times [0, 1] \times \mathbb{R},
\end{aligned}
$$

where

$$
\varphi_n(x, y) = \begin{cases} \left(\frac{y}{x}\right)^2 & \text{when } x \ge c_{m,M};\\ \left(\frac{3}{c_{m,M}^4}x^2 - \frac{8}{c_{m,M}^3}x + \frac{6}{c_{m,M}^2}\right)y^2 & \text{when } x < c_{m,M}. \end{cases}
$$

Elsewhere in this Appendix except in the proof of this Lemma, f_n should refer to the sequence of *univariate* functions defined in the beginning of Appendix A.3.

Then proving Lemma 8 reduces to proving

$$
E\sqrt{n}\,|V(n, f_n)_1| \le C\cdot o(1), \tag{52}
$$

where we adopt the notation of Delattre and Jacod (1997) and define, for $n \ge N_{m,M}$,

$$
\begin{aligned}
V(n, f_n)_t &:= \sum_{i=1}^{\lfloor nt\rfloor}\left(\frac{\widetilde{S}_{t_i} - \widetilde{S}_{t_{i-1}}}{\widetilde{S}_{t_{i-1}}}\right)^2 - \sum_{i=1}^{\lfloor nt\rfloor}\left(\frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}}\right)^2\\
&\quad - \frac{\beta_n^2}{6}\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}\frac{1}{S_{t_{i-1}}^2}\\
&= \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} f_n\left(S_{t_{i-1}}, \{S_{t_{i-1}}/\alpha_n\}, \sqrt{n}(S_{t_i} - S_{t_{i-1}})\right).
\end{aligned}
$$

Moreover, define

$$
\begin{aligned}
\breve{f}_n(x, u, y) &:= \varphi_n(x, \beta_n\lfloor u + y/\beta_n\rfloor) - \varphi_n(x, y)\\
&\quad - \frac{\beta_n^2}{6}\varphi_n(x, 1), \text{ for } (x, u, v) \in \mathbb{R}^+ \times [0, 1] \times \mathbb{R}.
\end{aligned}
$$

$\varphi_n(x, y)$ satisfies the Hypothesis L_2 of Delattre and Jacod (1997) and f_n and $\breve{f}_n$ satisfy the Hypothesis K_2' of Delattre and Jacod (1997). The limiting functions of f_n and $\breve{f}_n$, i.e., f and $\breve{f}$, are all zero.

Then we can write $\sqrt{n}V(n, f_n)_t$ as follows

$$
\begin{aligned}
&\sqrt{n}V(n, f_n)_t\\
&= \underbrace{\sqrt{n}\int_0^t M\breve{f}_n(S_s)ds}_{V_{n,1}(t)} + \underbrace{\sqrt{n}\left(\int_0^t \left(Mf_n(S_s) - M\breve{f}_n(S_s)\right)ds\right)}_{V_{n,2}(t)}\\
&\quad + \underbrace{\sqrt{n}\left(V(n, f_n)_t - \int_0^t Mf_n(S_s)ds\right)}_{V_{n,3}(t)},
\end{aligned}
$$

where $Mf(x) := \int_0^1 mf(x, u)du$, $mf(x, u) := \int f(x, u, \sigma(x)xy)h(y)dy$ and $h(y)$ is the density function of standard Normal distribution. It remains to prove $E\sup_{0\le t\le 1}|V_{n,k}(t)| \le C\cdot o(1)$, for $k = 1, 2$ and 3. We deal with them one by one next.

First, we treat $V_{n,1}(t)$ as follows.

$$
\begin{aligned}
V_{n,1}(t) &= \sqrt{n}\int_0^t\int_0^1\int (\varphi_n(S_s, \beta_n\lfloor u + \sigma_s S_s y/\beta_n\rfloor)\\
&\quad - \varphi_n(S_s, \sigma_s S_s y) - \frac{\beta_n^2}{6}\varphi_n(S_s, 1)\Big) h(y)dyduds\\
&= \sqrt{n}\int_0^t \left(\int_0^1\int h(y)\varphi_n(S_s, \beta_n\lfloor u + \sigma_s S_s y/\beta_n\rfloor)dydu\right.
\end{aligned}
$$

$$- \sigma_s^2 - \frac{\beta_n^2}{6}\frac{1}{S_s^2}\Big) ds$$
$$= -\sqrt{n}\int_0^t \frac{\beta_n^2}{\pi^2}\frac{1}{S_s^2}\sum_{j=1}^{\infty}\frac{1}{j^2}\exp\left(-2\pi^2 j^2 \frac{\sigma_s^2 S_s^2}{\beta_n^2}\right) ds,$$

where the last equality follows from the equation (A.10) of Li and Mykland (2015). By $\beta_n = O(n^{-\varsigma})$ for some $\varsigma > 0$, and the boundedness of both S_s and σ_s, we proved

$$E \sup_{0\le t\le 0} |V_{n,1}(t)| \le C \cdot o(1).$$

Second, by the definitions of f_n, $\breve{f}_n$ and $Mf(x)$, we can rewrite $V_{n,2}(t)$ as

$$\sqrt{n}\left(\int_0^t \int_0^1 \int (\varphi_n(S_s - \alpha_n u, \beta_n\lfloor u + \sigma_s S_s y/\beta_n\rfloor)\right.$$
$$\left. - \varphi_n(S_s, \beta_n\lfloor u + \sigma_s S_s y/\beta_n\rfloor))\, h(y)dyduds\right)$$
$$= -\sqrt{n}\int_0^t \int_0^1 \int \alpha_n u \varphi_n'(S_s - \lambda\alpha_n u, \beta_n\lfloor u + \sigma_s S_s y/\beta_n\rfloor)$$
$$\times\ h(y)dyduds,$$

where $\lambda \in (0, 1)$ and $\varphi_n'(x, y) := \partial\varphi_n(x, y)/\partial x$. Hence, by the fact that $\varphi_n(x, y)$ satisfies the Hypothesis L_2 of Delattre and Jacod (1997), we proved

$$E \sup_{0\le t\le 0} |V_{n,2}(t)| \le C \cdot \beta_n = C \cdot o(1).$$

Third, we decompose $V_{n,3}(t)$ as follows

$$V_{n,3}(t) = \underbrace{\sqrt{n}\left(V(n, f_n)_t - U_t^n\right)}_{V^1_{n,3}(t)} + \underbrace{\sqrt{n}\left(U_t^n - \int_0^t Mf_n(S_s)ds\right)}_{V^2_{n,3}(t)},$$

where

$$U_t^n = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} Mf_n(S_{t_{i-1}}).$$

By (b) of Lemma 5.3 in Delattre and Jacod (1997), we have

$$E \sup_{0\le t\le 1} |V^2_{n,3}(t)| \le C/\sqrt{n} = C \cdot o(1). \tag{53}$$

We further decompose $V^1_{n,3}(t)$ as

$$V^1_{n,3}(t) = M_t^n(k) + \widetilde{U}_t^n + \sqrt{n}(\overline{U}_t^n - U_t^n - \widetilde{U}_t^n/\sqrt{n}) - H_t^n(k),$$

where

$$\overline{U}_t^n = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} M_n f_n(S_{t_{i-1}}),\quad \widetilde{U}_t^n = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} M\tilde{f}_n(S_{t_{i-1}}),$$

and we refer readers to equations (2.12), (8.1), (8.7) and (8.12) in Delattre and Jacod (1997) for the definitions of $\tilde{f}_n$, $M_n f_n$, $M_t^n(k)$ and $H_t^n(k)$ respectively. Since the limiting function of $\tilde{f}_n$, i.e. $\tilde{f}$, is zero in our setting and by (a) of the proof of Theorem 2.2 in Delattre and Jacod (1997), we have

$$E \sup_{0\le t\le 1} |\widetilde{U}_t^n + \sqrt{n}(\overline{U}_t^n - U_t^n - \widetilde{U}_t^n/\sqrt{n}) - H_t^n(k_n)| \le C \cdot o(1),$$

by setting $k_n = \lfloor n^{3/4}\rfloor$. It remains to show for the locally square-integrable martingale $M_t^n(k_n)$ that $E\sup_{0\le t\le 1}|M_t^n(k_n)|^2 \le C \cdot o(1)$. By Doob's L^2 inequality

$$E \sup_{0\le t\le 1} |M_t^n(k_n)|^2 \le C \cdot E|M_1^n(k_n)|^2 = C \cdot E\left(\langle M^n(k_n), M^n(k_n)\rangle_1\right)$$
$$= C \cdot E\left(\frac{1}{n}\sum_{i=1}^{n} E_{S_{t_{i-1}}}(\mu_1^n(k_n)^2)\right).$$

Denote $C_1^n = 1/n\sum_{i=1}^n E_{S_{t_{i-1}}}(\mu_1^n(k_n)^2) := 1/n\sum_{i=1}^n \delta^n(k_n, S_{t_{i-1}})$. We refer to equation (8.6) of Delattre and Jacod (1997) for the definition of $\mu_1^n(k_n)$. The proof of $E|C_1^n| \le C \cdot o(1)$ can be divided into the following 4 steps.

(A) By part (c) of the Proof of Theorem 2.2 in Delattre and Jacod (1997)

$$\left|C_1^n - \frac{1}{n}\sum_{i=1}^{n}\tilde{\delta}^n(S_{t_{i-1}}, \{S_{t_{i-1}}/\alpha_n\})\right| \le Cn^{-3/32},$$

where $\tilde{\delta}^n(x, u) = \delta_{f_{n,x}, f_{n,x}}(\sigma(x), \beta_n, u)$ and $f_{n,x}(u, y) = f_n(x, u, y)$. We refer to equation (2.7) of Delattre and Jacod (1997) for the definition of $\delta_{f_{n,x}, f_{n,x}}$.

(B) The Remark 9.1 following the Proof of Theorem 2.1 in Delattre and Jacod (1997) implies that

$$E\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{\delta}^n(S_{t_{i-1}}, \{S_{t_{i-1}}/\alpha_n\}) - \frac{1}{n}\sum_{i=1}^{n} M\tilde{\delta}^n(S_{t_{i-1}})\right| \le C \cdot n^{-1/6},$$

where by (2.2), (2.7) and (2.11) of Delattre and Jacod (1997), $M\tilde{\delta}^n(x) = \Delta(f_n, f_n)(x, \beta_n)$ which satisfies K_1' of Delattre and Jacod (1997) as implied by (7.6) and (7.9) of Delattre and Jacod (1997).

(C) Part (a) of Lemma 5.3 in Delattre and Jacod (1997) implies that

$$E\left|\frac{1}{n}\sum_{i=1}^{n} M\tilde{\delta}^n(S_{t_{i-1}}) - \int_0^1 M\tilde{\delta}^n(S_s)ds\right| \le C \cdot o(1).$$

(D) Lemma 7.3 of Delattre and Jacod (1997) implies that $\Delta(f_n, f_n)(x, \beta_n) \to \Delta(f, f)(x, 0) \equiv 0$ pointwise. We have

$$E\left|\int_0^1 M\tilde{\delta}^n(S_s)ds\right| = E\left|\int_0^1 M\tilde{\delta}^n(S_s)ds - \int_0^1 \Delta(f, f)(S_s, 0)ds\right|$$
$$\le E\int_0^1 \left|M\tilde{\delta}^n(S_s) - \Delta(f, f)(S_s, 0)\right| ds = o(1)$$

by the bounded convergence theorem. This holds independent of the grid of times used.

We proved $E|C_1^n| \le C\cdot o(1)$, hence, $E\sup_{0\le t\le 1}|M_t^n(k_n)|^2 \le C\cdot o(1)$ and $E\sup_{0\le t\le 1}|V^1_{n,3}(t)| \le C \cdot o(1)$. This together with (53) implies that

$$E \sup_{0\le t\le 1} |V_{n,3}(t)| \le C \cdot o(1).$$

To sum up, we proved (52), completing the proof of Lemma 8. ■

Now we are ready to prove Proposition 6.

Proof of Proposition 6

The proof is divided into the following three parts.

Part A: The case where Condition RA(b) holds

Because Condition RA(b) and (6) imply $(\bar{n}_l)^{1/2} \cdot \bar{n}_l\alpha_n^2 = o(1)$, it suffices to prove the joint stable convergence in law for

$$\sqrt{\bar{n}_l}\left([Y, Y]_1^{(\mathrm{avg}, l)} - \int_0^1 \sigma_t^2 dt - 2\bar{n}_l\gamma_n^2\right),\ l = 1, 2, \ldots, L.$$

The result of Proposition 6 then follows by realizing that $\bar{n}_l = n/K_l$ and $K_l = \mathcal{C}_l n^{\kappa} + O(1)$. Applying the same decomposition (see Eq. (48)) of the γ_n-dominating case for the generic sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$, we have that

$$\begin{aligned}
&\sqrt{\bar{n}_l}\left([Y,Y]_1^{(\text{avg},l)} - \int_0^1 \sigma_t^2 dt - 2\bar{n}_l\gamma_n^2\right)\\
&= \sqrt{\bar{n}_l}\left(\frac{1}{K_l}\sum_{k=0}^{K_l-1}[Y,Y]_1^{\mathcal{G}_{k,l}} - \int_0^1 \sigma_t^2 dt - 2\bar{n}_l\gamma_n^2\right)\\
&= \sqrt{\bar{n}_l}\left(\frac{1}{K_l}\sum_{k=0}^{K_l-1}[\epsilon,\epsilon]_1^{\mathcal{G}_{k,l}} - 2\bar{n}_l\gamma_n^2\right) + \sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{k=0}^{K_l-1}[f_n(X),\epsilon]_1^{\mathcal{G}_{k,l}}\\
&\quad + \sqrt{\bar{n}_l}\left(\frac{1}{K_l}\sum_{k=0}^{K_l-1}[f_n(X),f_n(X)]_1^{\mathcal{G}_{k,l}} - \int_0^1 \sigma_t^2 dt\right). \qquad (54)
\end{aligned}$$

We deal with the three terms in (54) one by one as follows.

(i):

Using the same decomposition as in Part (b) of the proof of Proposition 3 for the sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$, we can write the first term in (54) as follows

$$\begin{aligned}
&\sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{i=1}^{n}\left(\epsilon_{t_i}^2 - E(\epsilon_{t_i}^2|X)\right)\\
&\quad - \sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}\epsilon_{t_{i,-}}\epsilon_{t_i} + \sqrt{\bar{n}_l}\cdot\left(\frac{2}{K_l}\sum_{i=1}^{n}E(\epsilon_{t_i}^2|X) - 2\bar{n}_l\gamma_n^2\right)\\
&\quad - \sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\max\mathcal{G}_{k,l}}} + \sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\min\mathcal{G}_{k,l}}},
\end{aligned}$$

where $t_{i,-}$ is the time that precedes t_i on sub-grid $\mathcal{G}_{k,l}$, and $t_{\max\mathcal{G}_{k,l}}$ and $t_{\min\mathcal{G}_{k,l}}$ are the maximum and minimum times on sub-grid $\mathcal{G}_{k,l}$.

$$\begin{aligned}
&E\left\{\left.\left(\sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{i=1}^{n}\left(\epsilon_{t_i}^2 - E(\epsilon_{t_i}^2|X)\right)\right)^2\right| X\right\}\\
&= \frac{4\bar{n}_l}{K_l^2}\sum_{i=1}^{n}\left(G_n(X_{t_i}) - g_n(X_{t_i})^2\right)\\
&= \frac{4\bar{n}_l}{K_l^2}\cdot\left(\sum_{i=1}^{n}2\gamma_n^4 + o_p(n\gamma_n^4)\right) = O_p\left(\frac{\bar{n}_l n\gamma_n^4}{K_l^2}\right)\\
&= O_p\left(\frac{1}{K_l}\right) = o_p(1)
\end{aligned}$$

by the assumption of Proposition 6 and the calculations for $G_n(x)$ and $g_n(x)$ in the proof of Lemma 7 and the proof of Proposition 3.

$$\begin{aligned}
&E\left\{\left.\left(\sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}\epsilon_{t_{i,-}}\epsilon_{t_i}\right)^2\right| X\right\}\\
&= \frac{4\bar{n}_l}{K_l^2}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}g_n(X_{t_{i,-}})g_n(X_{t_i}) = O_p\left(\frac{\bar{n}_l n\gamma_n^4}{K_l^2}\right)\\
&= O_p\left(\frac{1}{K_l}\right) = o_p(1)
\end{aligned}$$

by the calculation for $g_n(x)$ and the assumption of Proposition 6.

$$\begin{aligned}
&\sqrt{\bar{n}_l}\cdot\left(\frac{2}{K_l}\sum_{i=1}^{n}E(\epsilon_{t_i}^2|X) - 2\bar{n}_l\gamma_n^2\right)\\
&= \sqrt{\bar{n}_l}\cdot\left(\frac{2}{K_l}\sum_{i=1}^{n}g_n(X_{t_i}) - 2\bar{n}_l\gamma_n^2\right)\\
&= O_p\left(\sqrt{\bar{n}_l}\cdot\frac{n\alpha_n\gamma_n}{K_l}\right) = o_p(1),
\end{aligned}$$

by the calculation for $g_n(\cdot)$ and the assumption of Proposition 6.

$$\begin{aligned}
E\left(\left.\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\max\mathcal{G}_{k,l}}}\right| X\right) &= \frac{\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}g_n\left(X_{t_{\max\mathcal{G}_{k,l}}}\right)\\
&= O_p\left(\sqrt{\bar{n}_l}\gamma_n^2\right) = o_p(1)
\end{aligned}$$

and by the same calculation

$$E\left(\left.\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\min\mathcal{G}_{k,l}}}\right| X\right) = o_p(1).$$

Thus we proved

$$\sqrt{\bar{n}_l}\left(\frac{1}{K_l}\sum_{k=0}^{K_l-1}[\epsilon,\epsilon]_1^{\mathcal{G}_{k,l}} - 2\bar{n}_l\gamma_n^2\right) = o_p(1).$$

(ii):

Applying the same calculation for the sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$, we can write the conditional (on X) second moment for the second term in (54) as follows

$$\begin{aligned}
&E\left\{\left.\left(\sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{k=0}^{K_l-1}[f_n(X),\epsilon]_1^{\mathcal{G}_{k,l}}\right)^2\right| X\right\}\\
&= \frac{4\bar{n}_l}{K_l^2}\sum_{k=0}^{K_l-1}E\left[\left.\left([f_n(X),\epsilon]_1^{\mathcal{G}_{k,l}}\right)^2\right| X\right]\\
&= \frac{4\bar{n}_l}{K_l^2}\sum_{k=0}^{K_l-1}\left\{\sum_{t_i\in\mathcal{G}_{k,l}}\left(\Delta^{K_l}f_n(X_{t_{i,-}}) - \Delta^{K_l}f_n(X_{t_i})\right)^2 g_n(X_{t_i})\right.\\
&\quad + \left[\Delta^{K_l}f_n\left(X_{t_{\max\mathcal{G}_{k,l},-}}\right)\right]^2 g_n\left(X_{t_{\max\mathcal{G}_{k,l}}}\right)\\
&\quad \left.+\left[\Delta^{K_l}f_n\left(X_{t_{\min\mathcal{G}_{k,l}}}\right)\right]^2 g_n\left(X_{t_{\min\mathcal{G}_{k,l}}}\right)\right\}\\
&\le \frac{4\bar{n}_l}{K_l^2}\sum_{k=0}^{K_l-1}\sup_{x\in[m,M]}g_n(x)\times 4\times[f_n(X),f_n(X)]_1^{\mathcal{G}_{k,l}}\\
&= \frac{16\bar{n}_l}{K_l}\sup_{x\in[m,M]}g_n(x)\times[f_n(X),f_n(X)]_1^{(\text{avg},l)}\\
&= O_p\left(\frac{\bar{n}_l\gamma_n^2}{K_l}\right) = O_p\left(\frac{1}{K_l}\right) = o_p(1),
\end{aligned}$$

where $t_{\max\mathcal{G}_{k,l},-}$ is the time that precedes $t_{\max\mathcal{G}_{k,l}}$ on sub-grid $\mathcal{G}_{k,l}$, and Δ^{K_l} denotes the K_lth order difference operator over the complete grid $\mathcal{G}$ or the first order difference operator over the sub-grid $\mathcal{G}_{k,l}$, e.g., $\Delta^{K_l}f_n(X_{t_{i,-}}) = f_n(X_{t_i}) - f_n(X_{t_{i,-}})$ with $t_i, t_{i,-} \in \mathcal{G}_{k,l}$. Hence, the second term in (54) is negligible.

(iii):

We again apply the same calculation for the sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$. The third term in (54) can be written as

$$\begin{aligned}
&\sqrt{\bar{n}_l}\left([X,X]_1^{(\text{avg},l)} - \int_0^1 \sigma_t^2 dt\right)\\
&\quad + \sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}(f_n'(\zeta_i))^2 - 1)(X_{t_i} - X_{t_{i,-}})^2.
\end{aligned}$$

Observing that

$$\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}(f_n'(\zeta_i)^2-1)(X_{t_i}-X_{t_{i,-}})^2$$
$$\le\sqrt{\bar{n}_l}\times\sup_{x\in[m,M]}|f_n'(x)^2-1|\times\frac{1}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}(X_{t_i}-X_{t_{i,-}})^2$$
$$\le C\sqrt{\bar{n}_l}\times\sup_{x\in[m,M]}|f_n'(x)-1|\times[X,X]_1^{(\text{avg},l)}$$
$$=O_p\left(\sqrt{\bar{n}_l}\cdot\frac{\alpha_n}{\gamma_n}\right)=o_p(1),$$

which together with (i) and (ii) implies that only $\sqrt{\bar{n}_l}\left([X,X]_1^{(\text{avg},l)}-\int_0^1\sigma_t^2dt\right)$ plays a role in the central limit theorem. We deal with this term in Part C of the proof.

Part B: The case where Condition RO(b) holds

Because Condition RO(b) and (6) imply $\bar{n}_l\gamma_n^2=o(1/(\bar{n}_1)^{1/2})$ and $\gamma_\kappa=0$, it suffices to prove the joint stable convergence in law for

$$\sqrt{\bar{n}_l}\left([Y,Y]_1^{(\text{avg},l)}-\int_0^1\sigma_t^2dt-\frac{\bar{n}_l\alpha_n^2}{6}\int_0^1\frac{1}{S_t^2}dt\right),\ l=1,2,\dots,L.$$

The result of Proposition 6 then follows by realizing that $\bar{n}_l=n/K_l$ and $K_l=\mathcal{C}_ln^\kappa+O(1)$. Applying the same decomposition (see Eq. (42)) of the α_n-dominating case for the sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$, we rewrite the above term as follows

$$\begin{aligned}&\sqrt{\bar{n}_l}\left([R,R]_1^{(\text{avg},l)}+2[\widetilde{Y},R]_1^{(\text{avg},l)}\right)\\&+\sqrt{\bar{n}_l}\left([\widetilde{Y},\widetilde{Y}]_1^{(\text{avg},l)}-[X,X]^{(\text{avg},l)}-\frac{\bar{n}_l\alpha_n^2}{6}\int_0^1\frac{1}{S_t^2}dt\right)\\&+\sqrt{\bar{n}_l}\left([X,X]_1^{(\text{avg},l)}-\int_0^1\sigma_t^2dt\right).\end{aligned}\tag{55}$$

From the proof of Proposition 2 for the sub-grid $\mathcal{G}_s$, the same calculation applies to each sub-grid $\mathcal{G}_{k,l}$. Hence, we have that $\sqrt{\bar{n}_l}[\widetilde{Y},R]_1^{(\text{avg},l)}$ dominates in the first term of (55). Moreover,

$$E\left|\sqrt{\bar{n}_l}[\widetilde{Y},R]_1^{(\text{avg},l)}\right|=E\left|\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}[\widetilde{Y},R]_1^{\mathcal{G}_{k,l}}\right|$$
$$\le\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}E\left|[\widetilde{Y},R]_1^{\mathcal{G}_{k,l}}\right|=o(1),$$

because

$$E\left(\sqrt{\bar{n}_l}[\widetilde{Y},R]_1^{\mathcal{G}_{k,l}}\right)^2$$
$$\le C\bar{n}_l\left(\bar{n}_l\gamma_n^2+\alpha_n^2\max\left\{\frac{\bar{n}_lc_n}{b_n},\frac{\bar{n}_l^{1/2}c_n^{1/2}}{b_n^{1/2}}\right\}+\frac{C_1\bar{n}_l\alpha_n^2}{c_n}e^{-\frac{C_2c_n^2}{e^{2M}}}\right)$$

by applying the same calculation of the proof for Proposition 2 to each sub-grid $\mathcal{G}_{k,l}$ with $c_n=\log(b_n)$, where $b_n=\alpha_n/\gamma_n$ as defined in Eq. (26). Thus the first term of (55) is $o_p(1)$.

As to the second term of (55). By Lemma 8, since the same result applies to each sub-grid, we have that

$$\sqrt{\bar{n}_l}\left([\widetilde{Y},\widetilde{Y}]_1^{(\text{avg},l)}-[X,X]^{(\text{avg},l)}-\frac{\bar{n}_l\alpha_n^2}{6}\int_0^1\frac{1}{S_t^2}dt\right)=o_p(1).$$

Thus again only the third term in (55), i.e. $\sqrt{\bar{n}_l}([X,X]_1^{(\text{avg},l)}-\int_0^1\sigma_t^2dt)$, plays a role in the central limit theorem. We treat this term in Part C.

Part C: The central limit theorem

This part is dedicated to dealing with the joint stable convergence in law of the following quantities

$$\sqrt{\bar{n}_l}\left([X,X]_1^{(\text{avg},l)}-\int_0^1\sigma_t^2dt\right),\ l=1,\dots,L.$$

For $1\le l_1<l_2\le L$, denote two end points of martingales $D_1^{(1)}:=[X,X]_1^{(\text{avg},l_1)}-\langle X,X\rangle_1$ and $D_1^{(2)}:=[X,X]_1^{(\text{avg},l_2)}-\langle X,X\rangle_1$.

Follow Subsection A.3 of Zhang et al. (2005),

$$\begin{aligned}D_1^{(1)}&=[X,X]_1^{(\text{all})}-\langle X,X\rangle_1\\&\quad+2\sum_{i=1}^{n-1}\left(\sum_{j=1}^{K_{l_1}-1}\left(1-\frac{j}{K_{l_1}}\right)\Delta X_{t_{i-j}}\right)\Delta X_{t_i}+O_p(K_{l_1}/n).\end{aligned}$$

Here we allow the process starting from a time earlier than $t=0$ to make $\Delta X_{t_{i-j}}$ meaningful for $j\ge i$. Otherwise the remainder term incorporates end effects similarly as in Subsection A.3 of Zhang et al. (2005). Following Jacod and Protter (1998) and Mykland and Zhang (2006), $[X,X]_1^{(\text{all})}-\langle X,X\rangle_1=O_p(1/\sqrt{n})$. Therefore, it follows that given $K_{l_i}\to\infty$ and $K_{l_i}/n\to0$ for $i=1,2$,

$$D_1^{(1)}=2\sum_{i=1}^{n-1}\left(\sum_{j=1}^{K_{l_1}}\left(1-\frac{j}{K_{l_1}}\right)\Delta X_{t_{i-j}}\right)\Delta X_{t_i}+o_p\left(\sqrt{K_{l_1}/n}\right)$$

and similarly,

$$D_1^{(2)}=2\sum_{i=1}^{n-1}\left(\sum_{j=1}^{K_{l_2}}\left(1-\frac{j}{K_{l_2}}\right)\Delta X_{t_{i-j}}\right)\Delta X_{t_i}+o_p\left(\sqrt{K_{l_2}/n}\right).$$

Denote

$$C_i^{(1)}=\sum_{j=1}^{K_{l_1}}\left(1-\frac{j}{K_{l_1}}\right)\Delta X_{t_{i-j}}\quad\text{and}\quad C_i^{(2)}=\sum_{j=1}^{K_{l_2}}\left(1-\frac{j}{K_{l_2}}\right)\Delta X_{t_{i-j}}.$$

Then

$$\begin{aligned}\langle D^{(1)},D^{(2)}\rangle_1&=4\sum_{i=1}^{n-1}C_i^{(1)}C_i^{(2)}\Delta\langle X,X\rangle_{t_i}+o_p\left(\sqrt{K_{l_1}K_{l_2}}/n\right)\\&=(I)+(II)+(III)+(IV)+o_p\left(\sqrt{K_{l_1}K_{l_2}}/n\right),\end{aligned}$$

where

$$(I)=4\sum_{i=1}^{n-1}\Delta\langle X,X\rangle_{t_i}\sum_{j=1}^{K_{l_1}}\left(1-\frac{j}{K_{l_1}}\right)\left(1-\frac{j}{K_{l_2}}\right)(\Delta X_{t_{i-j}})^2,$$

$$(II)=4\sum_{i=1}^{n-1}\Delta\langle X,X\rangle_{t_i}\sum_{1\le l<k\le K_{l_1}}\left(1-\frac{l}{K_{l_1}}\right)\left(1-\frac{k}{K_{l_2}}\right)\Delta X_{t_{i-l}}\Delta X_{t_{i-k}},$$

$$(III)=4\sum_{i=1}^{n-1}\Delta\langle X,X\rangle_{t_i}\sum_{1\le k<l\le K_{l_1}}\left(1-\frac{l}{K_{l_1}}\right)\left(1-\frac{k}{K_{l_2}}\right)$$
$$\times\ \Delta X_{t_{i-l}}\Delta X_{t_{i-k}},$$

$$(IV)=4\sum_{i=1}^{n-1}\Delta\langle X,X\rangle_{t_i}\left(\sum_{l=i-K_{l_1}}^{i-1}\left(1-\frac{i-l}{K_{l_1}}\right)^+\Delta X_{t_l}\right)$$
$$\times\left(\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}\right)$$

and $a^+=\max(a,0)$. We determine the order of (IV) first by considering the equation in Box II.

$$
\begin{aligned}
(IV)' &= 4\sum_{i=1}^{n-1}\sigma^2_{t_{i-K_{l_2}}}\Delta t_i\left(\sum_{l=i-K_{l_1}}^{i-1}\left(1-\frac{i-l}{K_{l_1}}\right)^+\Delta X_{t_l}\right)\left(\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}\right)\\
&= \underbrace{4\sum_{l=1-K_{l_1}}^{0}\Delta X_{t_l}\sum_{i=1}^{n-1}\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\sigma^2_{t_{i-K_{l_2}}}\Delta t_i\left(1-\frac{i-l}{K_{l_1}}\right)^+\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}}_{I}\\
&\quad+\underbrace{4\sum_{l=1}^{n-2}\Delta X_{t_l}\sum_{i=l+1}^{n-1}\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\sigma^2_{t_{i-K_{l_2}}}\Delta t_i\left(1-\frac{i-l}{K_{l_1}}\right)^+\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}}_{II}.
\end{aligned}
$$

Box II.

Notice that

$$
\begin{aligned}
II &= 4\sum_{l=1}^{n-2}\Delta X_{t_l}\sum_{i=l+1}^{n-1}\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\sigma^2_{t_{i-K_{l_2}}}\Delta t_i\left(1-\frac{i-l}{K_{l_1}}\right)^+\\
&\quad\times\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}\\
&= 4\sum_{l=1}^{n-2}\Delta X_{t_l}(\zeta_l^{(1)}+\zeta_l^{(2)}),
\end{aligned}
$$

where

$$
\begin{cases}
\zeta_l^{(1)} = \displaystyle\sum_{k=l+1-K_{l_2}}^{l-K_{l_1}}\sum_{i=l+1}^{k+K_{l_2}}\sigma^2_{t_{i-K_{l_2}}}\Delta t_i\left(1-\frac{i-l}{K_{l_1}}\right)^+\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k},\\
\zeta_l^{(2)} = \displaystyle\sum_{k=l-K_{l_1}+1}^{l-1}\sum_{i=k+K_{l_1}+1}^{k+K_{l_2}}\sigma^2_{t_{i-K_{l_2}}}\Delta t_i\left(1-\frac{i-l}{K_{l_1}}\right)^+\\
\quad\times\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}.
\end{cases}
$$

Applying the Burkholder–Davis–Gundy inequality twice, we have that

$$
\begin{aligned}
E(II^2) &\le C\cdot E\sum_l \Delta\langle X,X\rangle_{t_l}(\zeta_l^{(1)}+\zeta_l^{(2)})^2\\
&\le \frac{C}{n}\cdot E\sum_l\sum_k\left(\sum_i\sigma^2_{t_{i-K_2}}\Delta t_i\left(1-\frac{i-l}{K_1}\right)^+\right.\\
&\quad\left.\times\left(1-\frac{i-k}{K_2}\right)^+\right)^2\Delta\langle X,X\rangle_{t_k}\\
&\le \frac{C}{n^2}\cdot E\sum_l\sum_k\left(\sum_i\sigma^2_{t_{i-K_2}}\Delta t_i\left(1-\frac{i-l}{K_1}\right)^+\right.\\
&\quad\left.\times\left(1-\frac{i-k}{K_2}\right)^+\right)^2\\
&\le \frac{C}{n^4}\cdot nK_{l_2}(K_{l_2}-K_{l_1})^2 = o\left(\frac{K_{l_1}K_{l_2}}{n^2}\right).
\end{aligned}
$$

The same argument as for II leads to $E(I^2)\le o\left(K_{l_1}K_{l_2}/n^2\right)$. Hence, we proved

$$
(IV)' = o_p\left(\frac{\sqrt{K_{l_1}K_{l_2}}}{n}\right).
$$

Moreover, by Hölder's inequality,

$$
\begin{aligned}
E|(IV)-(IV)'| &\le \|\sup_{|t-s|\le(K_2+1)/n}|\sigma_t^2-\sigma_s^2|\|_2\\
&\quad\times\sup_i\left\|\left(\sum_{l=i-K_{l_1}}^{i-1}\left(1-\frac{i-l}{K_{l_1}}\right)^+\Delta X_{t_l}\right)\right.\\
&\quad\left.\times\left(\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}\right)\right\|_2\\
&= o_p(K_{l_1}/n) = o_p(\sqrt{K_{l_1}K_{l_2}}/n),
\end{aligned}
$$

because of the boundedness and continuity of σ_t^2 and because

$$
\begin{aligned}
&E\left[\left(\sum_{l=i-K_{l_1}}^{i-1}\left(1-\frac{i-l}{K_{l_1}}\right)^+\Delta X_{t_l}\right)\left(\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}\right)\right]^2\\
&\quad\le O_p(K_{l_2}^2/n^2)
\end{aligned}
$$

uniformly in i by the Burkholder–Davis–Gundy inequality. Therefore,

$$
(IV) = o_p\left(\frac{\sqrt{K_{l_1}K_{l_2}}}{n}\right).
$$

Arguments parallel to that for (IV) give that

$$
(II) = o_p\left(\frac{\sqrt{K_{l_1}K_{l_2}}}{n}\right)\quad\text{and}\quad(III) = o_p\left(\frac{\sqrt{K_{l_1}K_{l_2}}}{n}\right).
$$

As to (I), we have that

$$
\begin{aligned}
(I) &= \frac{4}{n}\sum_{i=1}^{n-1}\sigma_{t_i}^4\Delta t_i\sum_{j=1}^{K_{l_1}}\left(1-\frac{j}{K_{l_1}}\right)\left(1-\frac{j}{K_{l_2}}\right)+o_p\left(\sqrt{K_{l_1}K_{l_2}}/n\right)\\
&= \frac{1}{n}\cdot\frac{(K_{l_1}-1)(3K_{l_2}-K_{l_1}-1)}{6K_{l_2}}\left(4\sum_{i=1}^{n-1}\sigma_{t_i}^4\Delta t_i\right)\\
&\quad+o_p\left(\sqrt{K_{l_1}K_{l_2}}/n\right).
\end{aligned}
$$

To sum up,

$$
\begin{aligned}
\sqrt{\bar n_{l_1}}\sqrt{\bar n_{l_2}}\langle D^{(1)},D^{(2)}\rangle_1 &= \frac{n}{\sqrt{K_{l_1}K_{l_2}}}\langle D^{(1)},D^{(2)}\rangle_1\\
&\xrightarrow{p}\frac{2\sqrt{\mathcal{C}_{l_1}}(3\mathcal{C}_{l_2}-\mathcal{C}_{l_1})}{3\mathcal{C}_{l_2}^{3/2}}\int_0^1\sigma_t^4dt.
\end{aligned}
$$

As special cases, $\bar n_{l_i}\langle D^{(i)},D^{(i)}\rangle_1\xrightarrow{p}4/3\int_0^1\sigma_t^4dt$, for $i=1,2$.

With an application of the argument (A.32) of Zhang et al. (2005) and that $\bar{n}_l = n/K_l$ and $K_l = \mathcal{C}_l n^{\kappa} + O(1)$, the limit result of Proposition 6 then follows from that in either Theorem B.4 (p. 65–67) of Zhang (2001) or Theorem 2.28 of Mykland and Zhang (2012). ∎

Proof of Theorem 1

The proof of the Theorem is completed by noticing that when K_l's are chosen according to (6),

$$\omega_l = \frac{1}{L} - \frac{\frac{1}{L}\sum_{j=1}^{L}\frac{1}{\mathcal{C}_j}\left(\frac{1}{\mathcal{C}_l} - \frac{1}{L}\sum_{j=1}^{L}\frac{1}{\mathcal{C}_j}\right)}{\sum_{j=1}^{L}\frac{1}{\mathcal{C}_j^2} - L\left(\frac{1}{L}\sum_{j=1}^{L}\frac{1}{\mathcal{C}_j}\right)^2} + o(1)$$
$$= \breve{\omega}_l + o(1) \quad \text{as } n \to \infty,$$

and hence the asymptotic variance of UV estimator in (15) multiplied by n/K_1

$$\frac{n}{K_1 n^{1-\kappa}} \sum_{l_1=1}^{L}\sum_{l_2=1}^{L} \omega_{l_1}\omega_{l_2}\tilde{q}_{l_1,l_2} \to \widetilde{\mathcal{V}} \quad \text{as } n \to \infty. \quad ∎$$

Proof of Proposition 7

We prove results over the complete grid. The case of sub-grid with averaging follows immediately and is treated in Part C.

Part A: Random noise dominates

In this case, $(2n)/3[Y,Y,Y,Y]_1^{(\mathrm{all})}$ can be rewritten and decomposed as follows

$$\begin{aligned}\frac{2n}{3}\sum_{i=0}^{n-1}\left(\Delta f_n(X_{t_i}) + \Delta\epsilon_{t_i}\right)^4 &= \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta f_n(X_{t_i}))^4 \\ &+ \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta\epsilon_{t_i})^4 + \frac{8n}{3}\sum_{i=0}^{n-1}(\Delta f_n(X_{t_i}))^3\Delta\epsilon_{t_i} \\ &+ \frac{12n}{3}\sum_{i=0}^{n-1}(\Delta f_n(X_{t_i}))^2(\Delta\epsilon_{t_i})^2 + \frac{8n}{3}\sum_{i=0}^{n-1}\Delta f_n(X_{t_i})(\Delta\epsilon_{t_i})^3. \end{aligned} \quad (56)$$

The terms on the right hand side of the above equation are treated one by one next.

(i):

The first term on the right hand side of (56) can be further decomposed as follows

$$\begin{aligned}&\frac{2n}{3}\sum_{i=0}^{n-1}(\Delta f_n(X_{t_i}))^4 \\ &= \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta f_n(X_{t_i}))^4 - \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^4 + \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^4 \\ &= \frac{2n}{3}\sum_{i=0}^{n-1}(f_n'(\zeta_i)^4 - 1)(\Delta X_{t_i})^4 + \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^4 \xrightarrow{p} 2\int_0^1 \sigma_t^4 dt,\end{aligned}$$

where the convergence is due to

$$\begin{aligned}E\left|\frac{2n}{3}\sum_{i=0}^{n-1}(f_n'(\zeta_i)^4 - 1)(\Delta X_{t_i})^4\right| &\le Cn \sup_{x\in[m,M]}|f_n'(x) - 1| \cdot E\sum_{i=0}^{n-1}(\Delta X_{t_i})^4 \\ &\le C\frac{\alpha_n}{\gamma_n} \to 0,\end{aligned}$$

by Lemma 3, where ζ_i is between X_{t_i} and $X_{t_{i+1}}$; and

$$\frac{2n}{3}[X,X,X,X]_1^{(\mathrm{all})} = \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^4 \xrightarrow{p} 2\int_0^1 \sigma_t^4 dt.$$

from Mykland and Zhang (2012).

(ii):

The expectation of the second term on the right hand side of (56) can be bounded as follows

$$\begin{aligned}E\frac{2n}{3}\sum_{i=0}^{n-1}(\Delta\epsilon_{t_i})^4 &\le Cn \cdot E\sum_{i=0}^{n-1}(\epsilon_{t_i}^4 + \epsilon_{t_{i+1}}^4) \le Cn^2 \sup_{x\in[m,M]} G_n(x) \\ &\le Cn^2\gamma_n^4 \to 0,\end{aligned}$$

by the calculation for $G_n(x)$ in the proof of Lemma 7.

(iii):

The second moment of the third term on the right hand side of (56) can be written and bounded as follows

$$\begin{aligned}&\frac{64n^2}{9}E\left(\sum_{i=1}^{n-1}\left[(\Delta f_n(X_{t_{i-1}}))^3 - (\Delta f_n(X_{t_i}))^3\right]\epsilon_{t_i}\right. \\ &\qquad \left. - (\Delta f_n(X_{t_0}))^3\epsilon_{t_0} + (\Delta f_n(X_{t_{n-1}}))^3\epsilon_{t_n}\right)^2 \\ &\le Cn^2 \sup_{x\in[m,M]} g_n(x) \cdot E\left(\sum_{i=1}^{n-1}\left[(\Delta f_n(X_{t_{i-1}}))^3 - (\Delta f_n(X_{t_i}))^3\right]^2\right. \\ &\qquad \left. + (\Delta f_n(X_{t_0}))^6 + (\Delta f_n(X_{t_{n-1}}))^6\right) \\ &\le Cn^2\gamma_n^2 \cdot n\frac{1}{n^3} = O(\gamma_n^2) \to 0\end{aligned}$$

by $E(\Delta f_n(X_{t_i}))^6 \le (\sup_{x\in[m,M]} f_n'(x))^6 \cdot E(\Delta X_{t_i})^6 \le C/n^3$ and the calculation for $g_n(x)$ in the proof of Proposition 3.

(iv):

The expectation of the fourth term on the right hand side of (56) can be bounded as follows

$$\begin{aligned}E\frac{12n}{3}\sum_{i=0}^{n-1}(\Delta f_n(X_{t_i}))^2(\Delta\epsilon_{t_i})^2 &\le Cn \sup_{x\in[m,M]} g_n(x) \cdot E\sum_{i=0}^{n-1}(\Delta f_n(X_{t_i}))^2 \\ &\le Cn\gamma_n^2 \to 0\end{aligned}$$

by the calculation for $g_n(x)$ and $E(\Delta f_n(X_{t_i}))^2 \le C/n$.

(v):

By the same argument as above, the expectation of the absolute value of the fifth term on the right hand side of (56) can be bounded as follows

$$E\left|\frac{8n}{3}\sum_{i=0}^{n-1}\Delta f_n(X_{t_i})(\Delta\epsilon_{t_i})^3\right| \le Cn^2\frac{1}{\sqrt{n}}\gamma_n^3 = C(\sqrt{n}\gamma_n)^3 \to 0.$$

Part B: Rounding effect dominates

In this case, $(2n)/3[Y,Y,Y,Y]_1^{(\mathrm{all})}$ can be rewritten and decomposed as follows

$$\begin{aligned}\frac{2n}{3}\sum_{i=0}^{n-1}\left(\Delta\widetilde{Y}_{t_i} + \Delta R_{t_i}\right)^4 &= \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta\widetilde{Y}_{t_i})^4 + \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta R_{t_i})^4 \\ &+ \frac{12n}{3}\sum_{i=0}^{n-1}(\Delta\widetilde{Y}_{t_i})^2(\Delta R_{t_i})^2 \\ &+ \frac{8n}{3}\sum_{i=0}^{n-1}(\Delta\widetilde{Y}_{t_i})^3\Delta R_{t_i} + \frac{8n}{3}\sum_{i=0}^{n-1}\Delta\widetilde{Y}_{t_i}(\Delta R_{t_i})^3.\end{aligned} \quad (57)$$

The terms on the right hand side of the above equation are treated one by one next.

(i):

Denote $D_{t_i} = \widetilde{Y}_{t_i} - X_{t_i}$. Then

$$|D_{t_i}| = \left|\log\left(\frac{(e^{X_{t_i}})^{(\alpha_n)}}{e^{X_{t_i}}}\right)\right| \le C\alpha_n$$

uniformly since $X_t \in [m, M]$ for all $t \in [0, 1]$. The first term on the right hand side of (57) can be rewritten and further decomposed as follows

$$\frac{2n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i}+\Delta D_{t_i})^4 = \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^4 + \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta D_{t_i})^4 + \frac{12n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^2(\Delta D_{t_i})^2 + \frac{8n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^3\Delta D_{t_i} + \frac{8n}{3}\sum_{i=0}^{n-1}\Delta X_{t_i}(\Delta D_{t_i})^3. \quad (58)$$

The first term on the right hand side of (58) has the following convergence

$$\frac{2n}{3}[X,X,X,X]_1^{(\text{all})} = \frac{2n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^4 \xrightarrow{p} 2\int_0^1 \sigma_t^4 dt,$$

by Mykland and Zhang (2012). The rest terms on the right hand side of (58) are negligible and analyzed one by one below.

$$\frac{2n}{3}\sum_{i=0}^{n-1}(\Delta D_{t_i})^4 \le Cn^2\alpha_n^4 \to 0.$$

$$E\frac{12n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^2(\Delta D_{t_i})^2 \le Cn\alpha_n^2 E\sum_{i=0}^{n-1}(\Delta X_{t_i})^2 \le Cn\alpha_n^2 \to 0.$$

$$E\left|\frac{8n}{3}\sum_{i=0}^{n-1}(\Delta X_{t_i})^3\Delta D_{t_i}\right| \le Cn\alpha_n\cdot E\left|\sum_{i=0}^{n-1}(\Delta X_{t_i})^3\right| \le C\sqrt{n}\alpha_n \to 0.$$

$$E\left|\frac{8n}{3}\sum_{i=0}^{n-1}\Delta X_{t_i}(\Delta D_{t_i})^3\right| \le Cn\alpha_n^3\cdot E\left|\sum_{i=0}^{n-1}\Delta X_{t_i}\right| \le C(\sqrt{n}\alpha_n)^3 \to 0.$$

(ii):
The expectation of the second term on the right hand side of (57) can be bounded as follows

$$E\frac{2n}{3}\sum_{i=0}^{n-1}(\Delta R_{t_i})^4 \le Cn\cdot E\sum_{i=1}^{n}R_{t_i}^4 \le Cn\cdot o(n\alpha_n^4) = o(n^2\alpha_n^4) \to 0$$

where the second inequality follows by the same argument as that used in the proof of Lemma 2.
(iii):
The expectation of the third term on the right hand side of (57) can be bounded as follows

$$E\frac{12n}{3}\sum_{i=0}^{n-1}(\Delta\widetilde{Y}_{t_i})^2(\Delta R_{t_i})^2 \le Cn\cdot\sum_{i=0}^{n-1}\left(E(\Delta\widetilde{Y}_{t_i})^4E(\Delta R_{t_i})^4\right)^{1/2} \le Cn\alpha_n^2 \to 0$$

where we used $E(\Delta R_{t_i})^4 \le C\alpha_n^4$ that follows from Part B(ii) and $E(\Delta\widetilde{Y}_{t_i})^4 \le C/n^2$ due to the assumption $\sqrt{n}\alpha_n \to 0$, $|D_{t_i}| \le C\alpha_n$ and $E(\Delta X_{t_i})^4 \le C/n^2$.
(iv):
The expectation of the absolute value of the fourth term on the right hand side of (57) can be bounded as follows

$$E\left|\frac{8n}{3}\sum_{i=0}^{n-1}(\Delta\widetilde{Y}_{t_i})^3\Delta R_{t_i}\right| \le Cn\cdot\sum_{i=0}^{n-1}\left(E(\Delta\widetilde{Y}_{t_i})^6E(\Delta R_{t_i})^2\right)^{1/2} \le C\sqrt{n}\alpha_n \to 0$$

by the same argument as that used in (iii).

(v):
The expectation of the absolute value of the fifth term on the right hand side of (57) can be bounded as follows

$$E\left|\frac{8n}{3}\sum_{i=0}^{n-1}\Delta\widetilde{Y}_{t_i}(\Delta R_{t_i})^3\right| \le Cn\cdot\sum_{i=0}^{n-1}\left(E(\Delta\widetilde{Y}_{t_i})^2E(\Delta R_{t_i})^6\right)^{1/2} \le C(\sqrt{n}\alpha_n)^3 \to 0$$

again by the same argument as that used in (iii).

Part C: Summary

To sum up, in either case, we showed that $2n/3[Y,Y,Y,Y]_1^{(\text{all})} \xrightarrow{p} 2\int_0^1\sigma_t^4dt$. In fact, it is the term $2n/3[X,X,X,X]_1^{(\text{all})}$ that converges to $2\int_0^1\sigma_t^4dt$. The L^1 norm $\|\cdot\|_1$ of all other terms in the decompositions of Parts A and B can be bounded by a universal $o(1)$ that is independent of the grid of times used. From Eqs. (108)–(144) of Mykland and Zhang (2012), the L^1 norm of $2n/3[X,X,X,X]_1^{(\text{all})} - 2\int_0^1\sigma_t^4dt$ can also be bounded by a universal $o(1)$ that is independent of the grid used. This last two points are true because of the boundedness of $(\sigma_t)_{t\ge0}$ and $(X_t)_{t\ge0}$ we assume in Appendix A.1. In view of the above discussion, we can apply the same calculation over the complete grid to each sub-grid and get the limit result under the assumptions of sub-grid case. ■

Proof of Corollary 1

Given $K_L/K = o(1)$ under Condition RA(b) or $K_L/K = O(1)$ under Condition RO(b), $\mathcal{Q}/2 \xrightarrow{p} \int_0^1\sigma_t^4dt$ by Proposition 7. Recall $\widehat{\widetilde{\mathcal{V}}}$ as defined in (18) and $h_{l_1,l_2} = 2\mathcal{C}_{l_1}(3\mathcal{C}_{l_2}-\mathcal{C}_{l_1})/(3\mathcal{C}_{l_2})$ for $1 \le l_1 \le l_2 \le L$, and $h_{l_2,l_1} = h_{l_1,l_2}$. From the proof of Theorem 1,

$$\frac{1}{\mathcal{C}_1}\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\omega_{l_1}\omega_{l_2}h_{l_1,l_2} \to \frac{1}{\mathcal{C}_1}\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\breve{\omega}_{l_1}\breve{\omega}_{l_2}h_{l_1,l_2} \quad \text{as } n\to\infty.$$

Therefore,

$$\widehat{\widetilde{\mathcal{V}}} = \left(\frac{1}{\mathcal{C}_1}\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\omega_{l_1}\omega_{l_2}h_{l_1,l_2}\right)\frac{\mathcal{Q}}{2} \xrightarrow{p} \left(\frac{1}{\mathcal{C}_1}\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\breve{\omega}_{l_1}\breve{\omega}_{l_2}h_{l_1,l_2}\right)\int_0^1\sigma_t^4dt = \widetilde{\mathcal{V}} \quad \text{as } n\to\infty$$

by noticing that $h_{l_1,l_2}\int_0^1\sigma_t^4dt = \tilde{q}_{l_1,l_2}$ as in (9). Hence, the Corollary is a result of Theorem 1 because the convergence is stable in distribution. ■

We need the following Lemma 9 before we prove Proposition 8. Let

$$M(l)_1 := \sqrt{\bar{n}_l}\left([X,X]_1^{(\text{avg},l)} - \int_0^1\sigma_t^2dt\right),$$

$$M_1^{(1,l)} := \sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{i=1}^{n}\left(\epsilon_{t_i}^2 - E(\epsilon_{t_i}^2|X)\right),$$

$$M_1^{(2,l)} := \sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}\epsilon_{t_{i,-}}\epsilon_{t_i},$$

and

$$M_1^{(3,l)} := \frac{2\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}\left(\Delta^{K_l}f_n(X_{t_{i,-}}) - \Delta^{K_l}f_n(X_{t_i})\right)\epsilon_{t_i},$$

where $t_{i,-}$ is the time that precedes t_i on sub-grid $\mathcal{G}_{k,l}$, Δ^{K_l} denotes the K_lth order difference operator over the complete grid

$\mathcal{G}$ or the first order difference operator over the sub-grid $\mathcal{G}_{k,l}$, e.g., $\Delta^{K_l} f_n(X_{t_{i,-}}) = f_n(X_{t_i}) - f_n(X_{t_{i,-}})$ with $t_i, t_{i,-} \in \mathcal{G}_{k,l}$.

Lemma 9. *Under the conditions of Proposition 8, we have that* $\langle M^{(2,l_1)}, M^{(3,l_2)}\rangle_1 = o_p(1)$,

$\langle M^{(2,l_1)}, M^{(2,l_2)}\rangle_1 = o_p(1)$ for $l_1 < l_2$,

$$\langle M^{(2,l)}, M^{(2,l)}\rangle_1 \xrightarrow{p} \frac{1}{K_l^{1/2}K_l^{1/2}} \cdot \frac{4\gamma^4}{K_l K_l},$$

$$\langle M^{(1,l_1)}, M^{(1,l_2)}\rangle_1 \xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}} \cdot \frac{8\gamma^4}{K_{l_1} K_{l_2}},$$

$$\langle M^{(3,l_1)}, M^{(3,l_2)}\rangle_1 \xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}} \cdot \frac{8\gamma^2}{K_{l_2}} \int_0^1 \sigma_t^2 dt,$$

and

$$\langle M(l_1), M(l_2)\rangle_1 \xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}} \left(\int_0^1 2\sigma_t^4 dt\right) \times \left(1 + \frac{(K_{l_1}-1)(3K_{l_2}-K_{l_1}-1)}{3K_{l_2}}\right).$$

Proof. For $l_1 \le l_2$, by the calculation of $g_n(x)$ in the proof of Proposition 3,

$$\begin{aligned} &E\left[\left(\langle M^{(2,l_1)}, M^{(3,l_2)}\rangle_1\right)^2 \Big| X\right] \\ &= \frac{16n^2}{K_{l_1}^3 K_{l_2}^3} \sum_i \left(\Delta^{K_{l_2}} f_n\left(X_{t_{i,-}^{l_2}}\right) - \Delta^{K_{l_2}} f_n\left(X_{t_i}\right)\right)^2 g_n(X_{t_i})^2 g_n\left(X_{t_{i,-}^{l_1}}\right) \\ &= \frac{16n^2}{K_{l_1}^3 K_{l_2}^3} \sum_i \left(\Delta^{K_{l_2}} f_n\left(X_{t_{i,-}^{l_2}}\right) - \Delta^{K_{l_2}} f_n\left(X_{t_i}\right)\right)^2 \left(\gamma_n^6 + o_p(\gamma_n^6)\right) \\ &= O_p(n^2\gamma_n^6) = o_p(1), \end{aligned}$$

noticing that

$$\sum_i \left(\Delta^{K_{l_2}} f_n\left(X_{t_{i,-}^{l_2}}\right) - \Delta^{K_{l_2}} f_n\left(X_{t_i}\right)\right)^2 \le 4[f_n(X), f_n(X)]_1^{(\text{avg},l_2)} = O_p(1),$$

where $\Delta^{K_{l_j}}$ denotes the K_{l_j}th difference operator on complete grid $\mathcal{G}$ or the first order difference operator over sub-grids $\mathcal{G}_{k,l_j}$ and $t_{i,-}^{l_j}$ are times on subgrids $\mathcal{G}_{k,l_j}$ that precede time t_i. We thus have that

$\langle M^{(2,l_1)}, M^{(3,l_2)}\rangle_1 = o_p(1)$ for $l_1 \le l_2$.

By similar calculation,

$\langle M^{(2,l_2)}, M^{(3,l_1)}\rangle_1 = o_p(1)$ for $l_1 \le l_2$.

Moreover, when $l_1 < l_2$

$$\langle M^{(2,l_1)}, M^{(2,l_2)}\rangle_1 = \frac{4n}{K_{l_1}^{3/2}K_{l_2}^{3/2}} \sum_i \epsilon_{t_{i,-}^{l_1}} \epsilon_{t_{i,-}^{l_2}} g_n(X_{t_i}).$$

Then

$$\begin{aligned} &E\left(\left(\langle M^{(2,l_1)}, M^{(2,l_2)}\rangle_1\right)^2 |X\right) \\ &= \frac{16n^2}{K_{l_1}^3 K_{l_2}^3} \sum_i g_n\left(X_{t_{i,-}^{l_1}}\right) g_n\left(X_{t_{i,-}^{l_2}}\right) g_n(X_{t_i})^2 \\ &= \frac{16n^3\gamma_n^8}{K_{l_1}^3 K_{l_2}^3} + o_p(n^3\gamma_n^8) = o_p(1). \end{aligned}$$

Hence,

$\langle M^{(2,l_1)}, M^{(2,l_2)}\rangle_1 = o_p(1)$ for $l_1 < l_2$.

When $l_1 = l_2 = l$

$$\begin{aligned} \langle M^{(2,l)}, M^{(2,l)}\rangle_1 &= \frac{4n}{K_l^{3/2}K_l^{3/2}} \sum_i \epsilon_{t_{i,-}^l}^2 g_n(X_{t_i}) \\ &= \frac{4n}{K_l^{3/2}K_l^{3/2}} \sum_i \left(\epsilon_{t_{i,-}^l}^2 - g_n\left(X_{t_{i,-}^l}\right)\right) g_n(X_{t_i}) \\ &\quad + \frac{4n}{K_l^{3/2}K_l^{3/2}} \sum_i g_n\left(X_{t_{i,-}^l}\right) g_n(X_{t_i}) \\ &= \frac{4n^2\gamma_n^4}{K_l^{3/2}K_l^{3/2}} + o_p(n^2\gamma_n^4) \xrightarrow{p} \frac{1}{K_l^{1/2}K_l^{1/2}} \cdot \frac{4\gamma^4}{K_l K_l} \end{aligned}$$

by the calculation of $g_n(x)$ and that

$$\begin{aligned} &E\left[\left(\frac{4n}{K_l^{3/2}K_l^{3/2}} \sum_i \left(\epsilon_{t_{i,-}^l}^2 - g_n\left(X_{t_{i,-}^l}\right)\right) g_n(X_{t_i})\right)^2 \Bigg| X\right] \\ &= \frac{16n^2}{K_l^3 K_l^3} \sum_i \left(G_n\left(X_{t_{i,-}^l}\right) - g_n\left(X_{t_{i,-}^l}\right)^2\right) g_n(X_{t_i})^2 \\ &= \frac{32n^3\gamma_n^8}{K_l^3 K_l^3} + o_p(n^3\gamma_n^8) = o_p(n^4\gamma_n^8), \end{aligned}$$

which follows from the calculation of $G_n(x)$ and $g_n(x)$ in the proof of Lemma 7 and the proof of Proposition 3.

We are left to calculate for $1 \le l_1 \le l_2 \le L$, the following quadratic covariations

$\langle M^{(1,l_1)}, M^{(1,l_2)}\rangle_1$, $\langle M^{(3,l_1)}, M^{(3,l_2)}\rangle_1$ and $\langle M(l_1), M(l_2)\rangle_1$.

We deal with them one by one.

We treat $\langle M^{(1,l_1)}, M^{(1,l_2)}\rangle_1$ as follows

$$\begin{aligned} \langle M^{(1,l_1)}, M^{(1,l_2)}\rangle_1 &= \frac{4n}{K_{l_1}^{3/2}K_{l_2}^{3/2}} \sum_{i=1}^n \left[E\left(\epsilon_{t_i}^4|X\right) - (E(\epsilon_{t_i}^2|X))^2\right] \\ &= \frac{4n}{K_{l_1}^{3/2}K_{l_2}^{3/2}} \sum_{i=1}^n \left(G_n(X_{t_i}) - g_n(X_{t_i})^2\right) \\ &= \frac{8n^2\gamma_n^4}{K_{l_1}^{3/2}K_{l_2}^{3/2}} + o_p(n^2\gamma_n^4) \xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}} \cdot \frac{8\gamma^4}{K_{l_1}K_{l_2}}. \end{aligned}$$

We treat $\langle M^{(3,l_1)}, M^{(3,l_2)}\rangle_1$ as follows

$$\begin{aligned} &\langle M^{(3,l_1)}, M^{(3,l_2)}\rangle_1 \\ &= \frac{4n}{K_{l_1}^{3/2}K_{l_2}^{3/2}} \sum_i \left(\Delta^{K_{l_1}} f_n(X_{t_{i,-}^{l_1}}) - \Delta^{K_{l_1}} f_n(X_{t_i})\right) \\ &\quad \times \left(\Delta^{K_{l_2}} f_n(X_{t_{i,-}^{l_2}}) - \Delta^{K_{l_2}} f_n(X_{t_i})\right) g_n\left(X_{t_i}\right) \\ &= \frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}} \sum_i \Delta^{K_{l_1}} f_n(X_{t_{i,-}^{l_1}}) \Delta^{K_{l_2}} f_n(X_{t_{i,-}^{l_2}}) \\ &\quad + \frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}} \sum_i \Delta^{K_{l_1}} f_n(X_{t_i}) \Delta^{K_{l_2}} f_n(X_{t_i}) \\ &\quad - \frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}} \sum_i \Delta^{K_{l_1}} f_n(X_{t_{i,-}^{l_1}}) \Delta^{K_{l_2}} f_n(X_{t_i}) \\ &\quad - \frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}} \sum_i \Delta^{K_{l_2}} f_n(X_{t_{i,-}^{l_2}}) \Delta^{K_{l_1}} f_n(X_{t_i}) \\ &\quad + o_p(n\gamma_n^2), \end{aligned} \tag{59}$$

by the calculation of $g_n(x)$. Since K_l's are fixed positive integers, we consider the situation on each sub-grid $\mathcal{G}_{k,l_2}$. By the Burkholder–Davis–Gundy inequality, the boundedness of $f_n'(x)$ and σ_t and that

$\Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}}) = f'_n(\zeta_i)\Delta^{K_{l_1}} X_{t^{l_1}_{i,-}} = O_p((K_{l_1}/n)^{1/2})$ where ζ_i is in between $X_{t^{l_1}_{i,-}}$ and X_{t_i}

$$E\left(\sum_{t_i\in\mathcal{G}_{k,l_2}} \Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\Delta^{K_{l_2}} f_n(X_{t_i})\right)^2 \le E\left(\sum_{t_i\in\mathcal{G}_{k,l_2}} \left(\Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\right)^2 \int_{t_i}^{t^{l_2}_{i,+}} f'_n(X_t)\sigma_t^2 dt\right) \le C\frac{K_{l_1}}{n}.$$

Hence, we have that in (59)

$$\frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}}\sum_i \Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\Delta^{K_{l_2}} f_n(X_{t_i}) = O_p(\sqrt{n}\gamma_n^2) = o_p(1).$$

By the same argument as above, we also have that in (59)

$$\frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}}\sum_i \Delta^{K_{l_2}} f_n(X_{t^{l_2}_{i,-}})\Delta^{K_{l_1}} f_n(X_{t_i}) = o_p(1).$$

Then for calculating $\langle M^{(3,l_1)}, M^{(3,l_2)}\rangle_1$, we are left with the first two terms in (59)

$$\frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}}\sum_i \Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\Delta^{K_{l_2}} f_n(X_{t^{l_2}_{i,-}}) \text{ and}$$
$$\frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}}\sum_i \Delta^{K_{l_1}} f_n(X_{t_i})\Delta^{K_{l_2}} f_n(X_{t_i}).$$

We take the above first term for example while the limit of the second term follows from the same method. We also consider the situation on a specific sub-grid $\mathcal{G}_{k,l_2}$ and then do summation over these sub-grids giving the result.

$$\begin{aligned}
&\sum_{t_i\in\mathcal{G}_{k,l_2}} \Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\Delta^{K_{l_2}} f_n(X_{t^{l_2}_{i,-}})\\
&= \sum_{t_i\in\mathcal{G}_{k,l_2}} \left(f_n(X_{t^{l_1}_{i,-}}) - f_n(X_{t^{l_2}_{i,-}})\right)\Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\\
&\quad + \sum_{t_i\in\mathcal{G}_{k,l_2}} \left(\Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\right)^2\\
&= O_p(1/\sqrt{n}) + \sum_{t_i\in\mathcal{G}_{k,l_2}} \left(\Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\right)^2,
\end{aligned} \tag{60}$$

because

$$\sum_{t_i\in\mathcal{G}_{k,l_2}} \left(f_n(X_{t^{l_1}_{i,-}}) - f_n(X_{t^{l_2}_{i,-}})\right)\Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}}) = O_p(1/\sqrt{n})$$

by the same argument as we treat the third and fourth terms in (59). Summation over sub-grids $\mathcal{G}_{k,l_2}$ of (60) gives

$$\begin{aligned}
&\frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}}\sum_i \Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\Delta^{K_{l_2}} f_n(X_{t^{l_2}_{i,-}})\\
&= O_p(1/\sqrt{n}) + \frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}}\sum_{k=1}^{K_{l_2}}\sum_{t_i\in\mathcal{G}_{k,l_2}} \left(\Delta^{K_{l_1}} f_n(X_{t^{l_1}_{i,-}})\right)^2\\
&= O_p(1/\sqrt{n}) + \frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}} K_{l_1}[f_n(X), f_n(X)]_1^{(\text{avg},l_1)}\\
&\xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\cdot\frac{4\gamma^2}{K_{l_2}}\int_0^1 \sigma_t^2 dt.
\end{aligned}$$

The same argument leads to that for the second term in (59)

$$\frac{4n\gamma_n^2}{K_{l_1}^{3/2}K_{l_2}^{3/2}}\sum_i \Delta^{K_{l_1}} f_n(X_{t_i})\Delta^{K_{l_2}} f_n(X_{t_i}) \xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\cdot\frac{4\gamma^2}{K_{l_2}}\int_0^1 \sigma_t^2 dt.$$

Thus we proved that

$$\langle M^{(3,l_1)}, M^{(3,l_2)}\rangle_1 \xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\cdot\frac{8\gamma^2}{K_{l_2}}\int_0^1 \sigma_t^2 dt.$$

We finally turn to $\langle M(l_1), M(l_2)\rangle_1$. Recall that

$$M(l_k)_1 = \sqrt{\bar{n}_{l_k}}\left([X,X]_1^{(\text{avg},l_k)} - \int_0^1 \sigma_t^2 dt\right) \text{ for } k = 1, 2.$$

By Subsection A.3 of Zhang et al. (2005), for $k = 1, 2$

$$\begin{aligned}
&M(l_k)_1\\
&= \sqrt{\bar{n}_{l_k}}\left([X,X]_1^{(\text{all})} - \int_0^1 \sigma_t^2 dt\right)\\
&\quad + 2\sqrt{\bar{n}_{l_k}}\sum_{i=1}^{n-1}\left(\sum_{j=1}^{K_{l_k}-1}\left(1-\frac{j}{K_{l_k}}\right)\Delta X_{t_{i-j}}\right)\Delta X_{t_i} + O_p(1/\sqrt{n}),
\end{aligned}$$

where we allow the process starting from a time earlier than $t = 0$ to make $\Delta X_{t_{i-j}}$ meaningful for $j \ge i$. Otherwise the remainder term incorporates end effects similarly as in Subsection A.3 of Zhang et al. (2005). We rewrite $M(l_k)_1$ as follows

$$M(l_k)_1 = \mathcal{M}_1^{(1,l_k)} + \mathcal{M}_1^{(2,l_k)} + o_p(1),$$

where

$$\mathcal{M}_1^{(1,l_k)} := 2\sqrt{\bar{n}_{l_k}}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}(X_t - X_{t_i})dX_t$$

and

$$\mathcal{M}_1^{(2,l_k)} := 2\sqrt{\bar{n}_{l_k}}\sum_{i=1}^{n-1} A(l_k)_i\Delta X_{t_i} \text{ and}$$

$$A(l_k)_i := \sum_{j=1}^{K_{l_k}-1}\left(1-\frac{j}{K_{l_k}}\right)\Delta X_{t_{i-j}}.$$

Then we can write

$$\langle M(l_1), M(l_2)\rangle_1 = I + II + III + IV + o_p(1).$$

Here,

$I := \langle \mathcal{M}^{(1,l_1)}, \mathcal{M}^{(1,l_2)}\rangle_1$, $II := \langle \mathcal{M}^{(1,l_1)}, \mathcal{M}^{(2,l_2)}\rangle_1$,
$III := \langle \mathcal{M}^{(2,l_1)}, \mathcal{M}^{(1,l_2)}\rangle_1$ and $IV := \langle \mathcal{M}^{(2,l_1)}, \mathcal{M}^{(2,l_2)}\rangle_1$.

$$I = \langle \mathcal{M}^{(1,l_1)}, \mathcal{M}^{(1,l_2)}\rangle_1 \xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\left(\int_0^1 2\sigma_t^4 dt\right)$$

following Mykland and Zhang (2006).

$$\begin{aligned}
II &= \langle \mathcal{M}^{(1,l_1)}, \mathcal{M}^{(2,l_2)}\rangle_1 = \frac{4n}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\sum_{i=1}^{n-1}\int_{t_i}^{t_{i+1}}\left(X_t - X_{t_i}\right)A(l_2)_i\sigma_t^2 dt\\
&= \frac{4n}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\sum_{i=1}^{n-1}\int_{t_i}^{t_{i+1}}(N_t - N_{t_i})\sigma_t^2 dt,
\end{aligned}$$

where

$$N_t := \sum_{t_{i+1}\le t} A(l_2)_i(X_{t_{i+1}} - X_{t_i}) + A(l_2)_{i^*}(X_t - X_{t_{i^*}})$$

and t_{i^*} is the largest $t_i \le t$. Since

$$\int_{t_i}^{t_{i+1}} (N_t - N_{t_i})\sigma_t^2 dt = \int_{t_i}^{t_{i+1}} (N_t - N_{t_i})\sigma_{t_i}^2 dt + \int_{t_i}^{t_{i+1}} (N_t - N_{t_i})(\sigma_t^2 - \sigma_{t_i}^2) dt$$

and by Ito's formula

$$d(t_{i+1} - t)(N_t - N_{t_i}) = -(N_t - N_{t_i})dt + (t_{i+1} - t)dN_t.$$

We can write II as follows

$$II = \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} (N_t - N_{t_i})\sigma_t^2 dt$$
$$= \underbrace{\frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} \sigma_{t_i}^2 \int_{t_i}^{t_{i+1}} (t_{i+1} - t) dN_t}_{II(1)} + \underbrace{\frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} (N_t - N_{t_i})(\sigma_t^2 - \sigma_{t_i}^2) dt}_{II(2)}.$$

By the Burkholder–Davis–Gundy inequality

$$E\left(II(1)^2\right) \le \frac{C16n^2}{K_{l_1}K_{l_2}} \cdot E \sum_{i=1}^{n-1} \sigma_{t_i}^4 \int_{t_i}^{t_{i+1}} (t_{i+1} - t)^2 A(l_2)_i^2 \sigma_t^2 dt$$
$$\le \frac{C16n^2}{K_{l_1}K_{l_2}} (t_{i+1} - t_i)^3 \cdot E \sum_{i=1}^{n-1} A(l_2)_i^2 \le \frac{C16n^2}{K_{l_1}K_{l_2}} \cdot \frac{1}{n^2} \frac{K_{l_2}}{n} = O(1/n)$$

because of the boundedness of σ_t and since by the Burkholder–Davis–Gundy inequality again that

$$EA(l_2)_i^2 \le C\frac{K_{l_2}}{n} \text{ uniformly in } i.$$

Thus

$$II(1) = O_p(1/\sqrt{n}) = o_p(1).$$

As to $II(2)$. We treat it as follows. Notice that by Cauchy–Schwarz inequality and Burkholder–Davis–Gundy inequality

$$\| (N_t - N_{t_i})(\sigma_t^2 - \sigma_{t_i}^2) \|_1 \le \| N_t - N_{t_i} \|_2 \, \| \sigma_t^2 - \sigma_{t_i}^2 \|_2$$
$$\le \left[C \cdot E\left([N,N]_t - [N,N]_{t_i}\right)\right]^{1/2} \| \sigma_t^2 - \sigma_{t_i}^2 \|_2$$
$$= \left[C \cdot E\left(\int_{t_i}^{t} A(l_2)_i^2 \sigma_s^2 ds\right)\right]^{1/2} \| \sigma_t^2 - \sigma_{t_i}^2 \|_2$$
$$\le \left[C\frac{K_{l_2}}{n^2}\right]^{1/2} \| \sigma_t^2 - \sigma_{t_i}^2 \|_2,$$

where the last inequality follows from again the fact that $EA(l_2)_i^2 \le CK_{l_2}/n$ uniformly in i. By the above result, we have that

$$E|II(2)| \le \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} \| (N_t - N_{t_i})(\sigma_t^2 - \sigma_{t_i}^2) \|_1 dt$$
$$\le \frac{4Cn}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \frac{K_{l_2}^{1/2}}{n} \sup_{0 \le t-s \le 1/n} \| \sigma_t^2 - \sigma_s^2 \|_2 = o(1)$$

by the continuity of σ_t. Therefore, we proved

$$II = o_p(1).$$

By the same method for proving $II = o_p(1)$, we can show that

$$III = o_p(1).$$

We next deal with IV.

$$IV = \langle \mathcal{M}^{(2,l_1)}, \mathcal{M}^{(2,l_2)} \rangle_1 = \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} A(l_1)_i A(l_2)_i \Delta\langle X, X \rangle_{t_i}$$
$$= \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} \left[\left(\sum_{j=1}^{K_{l_1}} \left(1 - \frac{j}{K_{l_1}}\right) \Delta X_{t_{i-j}} \right) \times \left(\sum_{j=1}^{K_{l_2}} \left(1 - \frac{j}{K_{l_2}}\right) \Delta X_{t_{i-j}} \right) \right] \Delta\langle X, X \rangle_{t_i}$$
$$= (I) + (II) + (III) + (IV),$$

where

$$(I) = \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} \Delta\langle X, X \rangle_{t_i} \times \sum_{j=1}^{K_{l_1}} \left(1 - \frac{j}{K_{l_1}}\right)\left(1 - \frac{j}{K_{l_2}}\right) (\Delta X_{t_{i-j}})^2,$$

$$(II) = \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} \Delta\langle X, X \rangle_{t_i} \times \sum_{1 \le l < k \le K_{l_1}} \left(1 - \frac{l}{K_{l_1}}\right)\left(1 - \frac{k}{K_{l_2}}\right) \Delta X_{t_{i-l}} \Delta X_{t_{i-k}},$$

$$(III) = \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} \Delta\langle X, X \rangle_{t_i} \times \sum_{1 \le k < l \le K_{l_1}} \left(1 - \frac{l}{K_{l_1}}\right)\left(1 - \frac{k}{K_{l_2}}\right) \Delta X_{t_{i-l}} \Delta X_{t_{i-k}},$$

$$(IV) = \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{i=1}^{n-1} \Delta\langle X, X \rangle_{t_i} \left(\sum_{l=i-K_{l_1}}^{i-1} \left(1 - \frac{i-l}{K_{l_1}}\right)^+ \Delta X_{t_l} \right) \times \left(\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1} \left(1 - \frac{i-k}{K_{l_2}}\right)^+ \Delta X_{t_k} \right)$$

and $a^+ = \max(a, 0)$. We determine the order of (IV) first by considering the equation in Box III.

Notice that

$$(ii) = \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{l=1}^{n-2} \Delta X_{t_l} \times \sum_{i=l+1}^{n-1} \sum_{k=i-K_{l_2}}^{i-K_{l_1}-1} \sigma_{t_{i-K_{l_2}}}^2 \Delta t_i \left(1 - \frac{i-l}{K_{l_1}}\right)^+ \left(1 - \frac{i-k}{K_{l_2}}\right)^+ \Delta X_{t_k}$$
$$= \frac{4n}{K_{l_1}^{1/2} K_{l_2}^{1/2}} \sum_{l=1}^{n-2} \Delta X_{t_l} (\zeta_l^{(1)} + \zeta_l^{(2)}),$$

where

$$\begin{cases} \zeta_l^{(1)} = \displaystyle\sum_{k=l+1-K_{l_2}}^{l-K_{l_1}} \sum_{i=l+1}^{k+K_{l_2}} \sigma_{t_{i-K_{l_2}}}^2 \Delta t_i \left(1 - \frac{i-l}{K_{l_1}}\right)^+ \left(1 - \frac{i-k}{K_{l_2}}\right)^+ \Delta X_{t_k}, \\ \zeta_l^{(2)} = \displaystyle\sum_{k=l-K_{l_1}+1}^{l-1} \sum_{i=k+K_{l_1}+1}^{k+K_{l_2}} \sigma_{t_{i-K_{l_2}}}^2 \Delta t_i \\ \quad \times \left(1 - \frac{i-l}{K_{l_1}}\right)^+ \left(1 - \frac{i-k}{K_{l_2}}\right)^+ \Delta X_{t_k}. \end{cases}$$

$$
\begin{aligned}
(IV)' &= \frac{4n}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\sum_{i=1}^{n-1}\sigma^2_{t_{i-K_{l_2}}}\Delta t_i\left(\sum_{l=i-K_{l_1}}^{i-1}\left(1-\frac{i-l}{K_{l_1}}\right)^+\Delta X_{t_l}\right)\left(\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}\right)\\
&= \underbrace{4\sum_{l=1-K_{l_1}}^{0}\Delta X_{t_l}\sum_{i=1}^{n-1}\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\sigma^2_{t_{i-K_{l_2}}}\Delta t_i\left(1-\frac{i-l}{K_{l_1}}\right)^+\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}}_{(i)}\\
&\quad+\underbrace{4\sum_{l=1}^{n-2}\Delta X_{t_l}\sum_{i=l+1}^{n-1}\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\sigma^2_{t_{i-K_{l_2}}}\Delta t_i\left(1-\frac{i-l}{K_{l_1}}\right)^+\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}}_{(ii)}.
\end{aligned}
$$

Box III.

Applying the Burkholder–Davis–Gundy inequality twice, we have that

$$
\begin{aligned}
E[(ii)^2] &\le C\cdot\frac{16n^2}{K_{l_1}K_{l_2}}E\sum_l \Delta\langle X,X\rangle_{t_l}(\zeta_l^{(1)}+\zeta_l^{(2)})^2\\
&\le \frac{16n^2}{K_{l_1}K_{l_2}}\frac{C}{n}\cdot E\sum_l\sum_k\left(\sum_i\sigma^2_{t_{i-K_2}}\Delta t_i\right.\\
&\quad\times\left.\left(1-\frac{i-l}{K_1}\right)^+\left(1-\frac{i-k}{K_2}\right)^+\right)^2\Delta\langle X,X\rangle_{t_k}\\
&\le \frac{16n^2}{K_{l_1}K_{l_2}}\frac{C}{n^2}\cdot E\sum_l\sum_k\left(\sum_i\sigma^2_{t_{i-K_2}}\Delta t_i\left(1-\frac{i-l}{K_1}\right)^+\right.\\
&\quad\times\left.\left(1-\frac{i-k}{K_2}\right)^+\right)^2\\
&\le \frac{16n^2}{K_{l_1}K_{l_2}}\frac{C}{n^4}\cdot nK_{l_2}(K_{l_2}-K_{l_1})^2 = O\left(\frac{1}{n}\right).
\end{aligned}
$$

By the same argument, we can show that $E[(i)^2]\le O(1/n)$. Hence, we proved

$$(IV)' = O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1).$$

Moreover, by Hölder's inequality,

$$
\begin{aligned}
E|(IV)-(IV)'| &\le \frac{4n}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\;\|\sup_{|t-s|\le(K_2+1)/n}|\sigma_t^2-\sigma_s^2|\|_2\\
&\quad\times\sup_i\left\|\left(\sum_{l=i-K_{l_1}}^{i-1}\left(1-\frac{i-l}{K_{l_1}}\right)^+\Delta X_{t_l}\right)\right.\\
&\quad\times\left.\left(\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}\right)\right\|_2\\
&= \frac{4n}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\times o_p(1/n) = o_p(1),
\end{aligned}
$$

because of the boundedness and continuity of σ_t^2 and because

$$
\begin{aligned}
&E\left[\left(\sum_{l=i-K_{l_1}}^{i-1}\left(1-\frac{i-l}{K_{l_1}}\right)^+\Delta X_{t_l}\right)\left(\sum_{k=i-K_{l_2}}^{i-K_{l_1}-1}\left(1-\frac{i-k}{K_{l_2}}\right)^+\Delta X_{t_k}\right)\right]^2\\
&\quad\le O_p(K_{l_2}^2/n^2)
\end{aligned}
$$

uniformly in i by the Burkholder–Davis–Gundy inequality. Therefore,

$$(IV) = o_p(1).$$

Arguments parallel to that for (IV) give that

$$(II) = o_p(1) \quad\text{and}\quad (III) = o_p(1).$$

As to (I), we have that

$$
\begin{aligned}
(I) &= \frac{4n}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\frac{1}{n}\sum_{i=1}^{n-1}\sigma^4_{t_i}\Delta t_i\sum_{j=1}^{K_{l_1}}\left(1-\frac{j}{K_{l_1}}\right)\left(1-\frac{j}{K_{l_2}}\right)+o_p(1)\\
&= \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\cdot\frac{(K_{l_1}-1)(3K_{l_2}-K_{l_1}-1)}{6K_{l_2}}\left(4\sum_{i=1}^{n-1}\sigma^4_{t_i}\Delta t_i\right)\\
&\quad+o_p(1).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
IV &= \langle\mathcal{M}^{(2,l_1)},\mathcal{M}^{(2,l_2)}\rangle_1\\
&\xrightarrow{p}\frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\cdot\frac{(K_{l_1}-1)(3K_{l_2}-K_{l_1}-1)}{3K_{l_2}}\left(2\int_0^1\sigma_t^4dt\right).
\end{aligned}
$$

Therefore, we proved

$$
\begin{aligned}
&\langle M(l_1),M(l_2)\rangle_1\\
&\xrightarrow{p}\frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\left(\int_0^1 2\sigma_t^4dt\right)\left(1+\frac{(K_{l_1}-1)(3K_{l_2}-K_{l_1}-1)}{3K_{l_2}}\right).\qquad\blacksquare
\end{aligned}
$$

Proof of Proposition 8

The proof is divided into the following three parts.

Part A: the case where Condition RA(b') holds

In this situation, since $n^{1/2}\cdot n\alpha_n^2 = o(1)$, it suffices to prove the joint stable convergence in law for

$$\sqrt{\bar n_l}\left([Y,Y]_1^{(\text{avg},l)}-\int_0^1\sigma_t^2dt-2\bar n_l\gamma_n^2\right),\; l=1,2,\ldots,L.$$

The result of Proposition 8 then follows by realizing that $\bar n_l = n/K_l$ and K_l's are *fixed* positive integers. Applying the same decomposition (see Eq. (48)) of the γ_n-dominating case for the sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$, we have that

$$\sqrt{\bar{n}_l}\left([Y,Y]_1^{(\mathrm{avg},l)}-\int_0^1\sigma_t^2dt-2\bar{n}_l\gamma_n^2\right)$$
$$=\sqrt{\bar{n}_l}\left(\frac{1}{K_l}\sum_{k=0}^{K_l-1}[Y,Y]_1^{\mathcal{G}_{k,l}}-\int_0^1\sigma_t^2dt-2\bar{n}_l\gamma_n^2\right)$$
$$=\sqrt{\bar{n}_l}\left\{\frac{1}{K_l}\sum_{k=0}^{K_l-1}\left([\epsilon,\epsilon]_1^{\mathcal{G}_{k,l}}+[f_n(X),f_n(X)]_1^{\mathcal{G}_{k,l}}+2[f_n(X),\epsilon]_1^{\mathcal{G}_{k,l}}\right)-\int_0^1\sigma_t^2dt-2\bar{n}_l\gamma_n^2\right\}$$
$$=\sqrt{\bar{n}_l}\left(\frac{1}{K_l}\sum_{k=0}^{K_l-1}[\epsilon,\epsilon]_1^{\mathcal{G}_{k,l}}-2\bar{n}_l\gamma_n^2\right)+\sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{k=0}^{K_l-1}[f_n(X),\epsilon]_1^{\mathcal{G}_{k,l}}$$
$$+\sqrt{\bar{n}_l}\left(\frac{1}{K_l}\sum_{k=0}^{K_l-1}[f_n(X),f_n(X)]_1^{\mathcal{G}_{k,l}}-\int_0^1\sigma_t^2dt\right). \tag{61}$$

We deal with the three terms in (61) one by one as follows.

(i):

Using the same decomposition as in part (b) of the proof of Proposition 3 for the sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$, we can write the first term in (61) as follows

$$\sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{i=1}^{n}\left(\epsilon_{t_i}^2-E(\epsilon_{t_i}^2|X)\right)-\sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}\epsilon_{t_{i,-}}\epsilon_{t_i}$$
$$+\sqrt{\bar{n}_l}\cdot\left(\frac{2}{K_l}\sum_{i=1}^{n}E(\epsilon_{t_i}^2|X)-2\bar{n}_l\gamma_n^2\right)$$
$$-\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\max\mathcal{G}_{k,l}}}+\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\min\mathcal{G}_{k,l}}}$$
$$=M_1^{(1,l)}-M_1^{(2,l)}+\sqrt{\bar{n}_l}\cdot\left(\frac{2}{K_l}\sum_{i=1}^{n}E(\epsilon_{t_i}^2|X)-2\bar{n}_l\gamma_n^2\right)$$
$$-\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\max\mathcal{G}_{k,l}}}$$
$$+\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\min\mathcal{G}_{k,l}}}, \tag{62}$$

where $M_1^{(1,l)}$ and $M_1^{(2,l)}$ are the end points of martingales that play roles in the central limit theorem, and $t_{\max\mathcal{G}_{k,l}}$ and $t_{\min\mathcal{G}_{k,l}}$ are the maximum and minimum times on sub-grid $\mathcal{G}_{k,l}$.

We deal with $M_1^{(1)}$ and $M_1^{(2)}$ in Part C and show that the rest terms in (62) are negligible below.

$$\sqrt{\bar{n}_l}\cdot\left(\frac{2}{K_l}\sum_{i=1}^{n}E(\epsilon_{t_i}^2|X)-2\bar{n}_l\gamma_n^2\right)$$
$$=\sqrt{\bar{n}_l}\cdot\left(\frac{2}{K_l}\sum_{i=1}^{n}g_n(X_{t_i})-2\bar{n}_l\gamma_n^2\right)$$
$$=O_p\left(\sqrt{\bar{n}_l}\cdot\frac{n\alpha_n\gamma_n}{K_l}\right)=o_p(1),$$

by the calculation of $g_n(\cdot)$ in the proof of Proposition 3 and the assumption of Proposition 8.

$$E\left(\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\max\mathcal{G}_{k,l}}}\middle|X\right)=\frac{\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}g_n\left(X_{t_{\max\mathcal{G}_{k,l}}}\right)$$
$$=O_p\left(\sqrt{\bar{n}_l}\gamma_n^2\right)=o_p(1)$$

and similarly,

$$E\left(\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\min\mathcal{G}_{k,l}}}\middle|X\right)=o_p(1)$$

by again the calculation of $g_n(\cdot)$ in the proof of Proposition 3 and the assumption of Proposition 8. Hence,

$$-\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\max\mathcal{G}_{k,l}}}+\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\epsilon^2_{t_{\min\mathcal{G}_{k,l}}}=o_p(1).$$

(ii):

Applying the same calculation for the sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$, we can write the second term in (61) as follows

$$\frac{2\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}\left(\Delta^{K_l}f_n(X_{t_{i,-}})-\Delta^{K_l}f_n(X_{t_i})\right)\epsilon_{t_i}$$
$$+\frac{2\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}\Delta^{K_l}f_n\left(X_{t_{\max\mathcal{G}_{k,l},-}}\right)\epsilon_{t_{\max\mathcal{G}_{k,l}}}$$
$$-\frac{2\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}\Delta^{K_l}f_n\left(X_{t_{\min\mathcal{G}_{k,l}}}\right)\epsilon_{t_{\min\mathcal{G}_{k,l}}}$$
$$=M_1^{(3,l)}+\frac{2\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}\Delta^{K_l}f_n\left(X_{t_{\max\mathcal{G}_{k,l},-}}\right)\epsilon_{t_{\max\mathcal{G}_{k,l}}}$$
$$-\frac{2\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}\Delta^{K_l}f_n\left(X_{t_{\min\mathcal{G}_{k,l}}}\right)\epsilon_{t_{\min\mathcal{G}_{k,l}}}, \tag{63}$$

where $M_1^{(3,l)}$ is the end point of a martingale that also plays a role in the central limit theorem, and $t_{\max\mathcal{G}_{k,l},-}$ is the time that precedes $t_{\max\mathcal{G}_{k,l}}$ on $\mathcal{G}_{k,l}$.

We deal with $M_1^{(3,l)}$ in Part C and show the rest terms in (63) are negligible below. Since $\Delta^{K_l}f_n(X_{t_i})=f_n'(\zeta_i)\Delta^{K_l}X_{t_i}=O_p(1/n^{1/2})$ by Lemma 3 where ζ_i lies in between X_{t_i} and $X_{t_{i,+}}$, $t_{i,+}$ and t_i are times on $\mathcal{G}_{k,l}$. We can easily see that $\sqrt{\bar{n}_l}\Delta^{K_l}f_n\left(X_{t_{\max\mathcal{G}_{k,l},-}}\right)\epsilon_{t_{\max\mathcal{G}_{k,l}}}=O_p(\gamma_n)$. Hence

$$\frac{2\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}\Delta^{K_l}f_n\left(X_{t_{\max\mathcal{G}_{k,l},-}}\right)\epsilon_{t_{\max\mathcal{G}_{k,l}}}$$
$$-\frac{2\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}\Delta^{K_l}f_n\left(X_{t_{\min\mathcal{G}_{k,l}}}\right)\epsilon_{t_{\min\mathcal{G}_{k,l}}}=O_p(\gamma_n)=o_p(1).$$

(iii):

Again, we apply the same calculation for the sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$. The third term in (61) can be written as

$$M(l)_1+\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}(f_n'(\zeta_i)^2-1)(X_{t_i}-X_{t_{i,-}})^2.$$

Observe that

$$\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}(f_n'(\zeta_i)^2-1)(X_{t_i}-X_{t_{i,-}})^2$$
$$\le\sqrt{\bar{n}_l}\times\sup_{x\in[m,M]}|f_n'(x)^2-1|\times\frac{1}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}(X_{t_i}-X_{t_{i,-}})^2$$
$$\le C\sqrt{\bar{n}_l}\times\sup_{x\in[m,M]}|f_n'(x)-1|\times[X,X]_1^{(\mathrm{avg},l)}=O_p\left(\frac{\sqrt{n}\alpha_n}{\gamma_n}\right)=o_p(1).$$

$M(l)_1$ is the end point of a martingale that plays a role in the central limit theorem. We deal with this term in Part C of the proof.

Part B: the case where Condition RO(b') holds
In this situation, since $n\gamma_n^2 = o(1/n^{1/2})$ and $\gamma = 0$ as implied by Condition RO(b'). It suffices to prove the joint stable convergence in law for

$$\sqrt{\bar{n}_l}\left([Y,Y]_1^{(\text{avg},l)} - \int_0^1 \sigma_t^2 dt - \frac{n\alpha_n^2}{6K_l}\int_0^1 \frac{1}{S_t^2}dt\right),\ l = 1,2,\ldots,L.$$

The result of Proposition 8 then follows by realizing that $\bar{n}_l = n/K_l$ and K_l's are *fixed* positive integers. Applying the same decomposition (see Eq. (42)) of the α_n-dominating case for the sub-grid $\mathcal{G}_s$ to each sub-grid $\mathcal{G}_{k,l}$, we have that

$$\begin{aligned}
&\sqrt{\bar{n}_l}\left([Y,Y]_1^{(\text{avg},l)} - \int_0^1 \sigma_t^2 dt - \frac{n\alpha_n^2}{6K_l}\int_0^1 \frac{1}{S_t^2}dt\right)\\
&= \sqrt{\bar{n}_l}\left([R,R]_1^{(\text{avg},l)} + 2[\widetilde{Y},R]_1^{(\text{avg},l)}\right)\\
&+ \sqrt{\bar{n}_l}\left([\widetilde{Y},\widetilde{Y}]_1^{(\text{avg},l)} - [X,X]^{(\text{avg},l)} - \frac{n\alpha_n^2}{6K_l}\int_0^1 \frac{1}{S_t^2}dt\right)\\
&+ \sqrt{\bar{n}_l}\left([X,X]_1^{(\text{avg},l)} - \int_0^1 \sigma_t^2 dt\right). \qquad (64)
\end{aligned}$$

From the proof of Proposition 2 for the sub-grid $\mathcal{G}_s$, the same calculation applies to each sub-grid $\mathcal{G}_{k,l}$. Hence, we have that $\sqrt{\bar{n}_l}[\widetilde{Y},R]_1^{(\text{avg},l)}$ dominates in the first term of (64). Moreover,

$$\begin{aligned}
E\left|\sqrt{\bar{n}_l}[\widetilde{Y},R]_1^{(\text{avg},l)}\right| &= E\left|\sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}[\widetilde{Y},R]_1^{\mathcal{G}_{k,l}}\right|\\
&\le \sqrt{\bar{n}_l}\cdot\frac{1}{K_l}\sum_{k=0}^{K_l-1}E\left|[\widetilde{Y},R]_1^{\mathcal{G}_{k,l}}\right| = o(1),
\end{aligned}$$

because

$$E\left(\sqrt{\bar{n}_l}[\widetilde{Y},R]_1^{\mathcal{G}_{k,l}}\right)^2 \le C\bar{n}_l\left(\bar{n}_l\gamma_n^2 + \alpha_n^2\max\left\{\frac{\bar{n}_l c_n}{b_n}, \frac{\bar{n}_l^{1/2}c_n^{1/2}}{b_n^{1/2}}\right\} + \frac{C_1\bar{n}_l\alpha_n^2}{c_n}e^{-\frac{C_2 c_n^2}{e^{2M}}}\right)$$

by applying the same calculation of the proof for Proposition 2 to each sub-grid $\mathcal{G}_{k,l}$ with $c_n = \log(b_n)$, where $b_n = \alpha_n/\gamma_n$ as defined in (26). Thus the first term of (64) is $o_p(1)$.

As to the second term of (64). By applying the central limit theorem of Li and Mykland (2015) to each sub-grid $\mathcal{G}_{k,l}$

$$\sqrt{\bar{n}_l}\left([\widetilde{Y},\widetilde{Y}]_1^{\mathcal{G}_{k,l}} - \int_0^1 \sigma_t^2 dt - \frac{n\alpha_n^2}{6K_l}\int_0^1 \frac{1}{S_t^2}dt\right) \to \left(\int_0^1 2\sigma_t^4 dt\right)^{1/2} N(0,1)$$

stably in law and noticing that $\sqrt{\bar{n}_l}([X,X]_1^{\mathcal{G}_{k,l}} - \int_0^1 \sigma_t^2 dt) \to \left(\int_0^1 2\sigma_t^4 dt\right)^{1/2} N(0,1)$ stably in law, we deduce that

$$\sqrt{\bar{n}_l}\left([\widetilde{Y},\widetilde{Y}]_1^{(\text{avg},l)} - [X,X]_1^{(\text{avg},l)} - \frac{n\alpha_n^2}{6K_l}\int_0^1 \frac{1}{S_t^2}dt\right) = o_p(1).$$

Thus only the third term in (64), i.e. $M(l)_1 = \sqrt{\bar{n}_l}\left([X,X]_1^{(\text{avg},l)} - \int_0^1 \sigma_t^2 dt\right)$, plays a role in the central limit theorem. We treat this term in Part C.

Part C: The central limit theorems
From Part A and B, we see that under Condition RA(b')

$$\begin{aligned}
&\sqrt{\bar{n}_l}\left([Y,Y]_1^{(\text{avg},l)} - \int_0^1 \sigma_t^2 dt - 2\frac{\gamma^2}{K_l}\right)\\
&= M_1^{(1,l)} + M_1^{(2,l)} + M_1^{(3,l)} + M(l)_1 + o_p(1),
\end{aligned}$$

where for clarity,

$$\begin{aligned}
M_1^{(1,l)} &:= \sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{i=1}^{n}\left(\epsilon_{t_i}^2 - E(\epsilon_{t_i}^2|X)\right),\\
M_1^{(2,l)} &:= \sqrt{\bar{n}_l}\cdot\frac{2}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}\epsilon_{t_{i,-}}\epsilon_{t_i},\\
M_1^{(3,l)} &:= \frac{2\sqrt{\bar{n}_l}}{K_l}\sum_{k=0}^{K_l-1}\sum_{t_i\in\mathcal{G}_{k,l}}\left(\Delta^{K_l}f_n(X_{t_{i,-}}) - \Delta^{K_l}f_n(X_{t_i})\right)\epsilon_{t_i} \text{ and}\\
M(l)_1 &:= \sqrt{\bar{n}_l}\left([X,X]_1^{(\text{avg},l)} - \int_0^1 \sigma_t^2 dt\right);
\end{aligned}$$

and under Condition RO(b')

$$\sqrt{\bar{n}_l}\left([Y,Y]_1^{(\text{avg},l)} - \int_0^1 \sigma_t^2 dt - \frac{n\alpha_n^2}{6K_l}\int_0^1 \frac{1}{S_t^2}dt\right) = M(l)_1 + o_p(1).$$

We only show the limit results under Condition RA(b') since limit results under Condition RO(b') are only part of that under Condition RA(b').

For $1 \le l_1 \le l_2 \le L$, we obviously have $\langle M^{(i,l_j)}, M(l_k)\rangle_1 = 0$ for $1 \le i \le 3$ and $1 \le j,k \le 2$; and $\langle M^{(1,l_i)}, M^{(k,l_j)}\rangle_1 = 0$ for $1 \le i,j \le 2$ and $k = 2,3$.

By Lemma 9, we have the following

$$\langle M^{(2,l_1)}, M^{(2,l_2)}\rangle_1 = o_p(1) \text{ for } l_1 < l_2,\ \langle M^{(2,l_1)}, M^{(3,l_2)}\rangle_1 = o_p(1),$$

$$\begin{aligned}
\langle M^{(2,l)}, M^{(2,l)}\rangle_1 &\xrightarrow{p} \frac{1}{K_l^{1/2}K_l^{1/2}}\cdot\frac{4\gamma^4}{K_l K_l},\\
\langle M^{(1,l_1)}, M^{(1,l_2)}\rangle_1 &\xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\cdot\frac{8\gamma^4}{K_{l_1}K_{l_2}},\\
\langle M^{(3,l_1)}, M^{(3,l_2)}\rangle_1 &\xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\cdot\frac{8\gamma^2}{K_{l_2}}\int_0^1 \sigma_t^2 dt,
\end{aligned}$$

and finally,

$$\begin{aligned}
&\langle M(l_1), M(l_2)\rangle_1\\
&\xrightarrow{p} \frac{1}{K_{l_1}^{1/2}K_{l_2}^{1/2}}\left(\int_0^1 2\sigma_t^4 dt\right)\left(1 + \frac{(K_{l_1}-1)(3K_{l_2}-K_{l_1}-1)}{3K_{l_2}}\right).
\end{aligned}$$

To sum up. The conditional Lyapunov conditions are satisfied by using the same yet tedious calculation as that for the previous quadratic (co)variations for $M^{(k,l)}$, $k = 1,2,3$ martingales. With an application of the argument (A.32) of Zhang et al. (2005), the limit results related to $M(l)$ martingale then follow from that in either Theorem B.4 (p. 65–67) of Zhang (2001) or Theorem 2.28 of Mykland and Zhang (2012). This together with a application of the martingale central limit theorem in view of Hall and Heyde (1998) (Corollary 3.1, p.58) and that $\bar{n}_l = n/K_l$ and K_l's are *fixed* positive integers confirms the limit result of Proposition 8. ■

Proof of Theorem 2
Theorem 2 is proved by noticing the asymptotic variance formula (21) and the fact that $\omega_l \equiv \widetilde{\omega}_l$ when K_l's are *fixed* positive integers. ■

Proof of Corollary 2

First, under Condition RO(b'), $\gamma = 0$, and under Condition RA(b'), $\widehat{\gamma} \xrightarrow{p} \gamma$ by Proposition 4 being applied to the complete grid (cf. Remark 4). Second, $\widehat{\langle X, X\rangle}_1 \xrightarrow{p} \int_0^1 \sigma_t^2 dt$ by Theorem 2. Third, given $K_L/K = o(1)$ under Condition RA(b') or $K_L/K = O(1)$ under Condition RO(b'), $\mathcal{Q}/2 \xrightarrow{p} \int_0^1 \sigma_t^4 dt$ by Proposition 7. Because ψ is continuous, the continuous mapping theorem leads to

$$\widehat{\mathcal{V}} = \psi\left(\widehat{\gamma}, \widehat{\langle X, X\rangle}_1, \frac{\mathcal{Q}}{2}\right) \xrightarrow{p} \psi\left(\gamma, \int_0^1 \sigma_t^2 dt, \int_0^1 \sigma_t^4 dt\right) = \mathcal{V}.$$

The Corollary is thus a result of Theorem 2 because the convergence is stable in distribution. ■

References

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financ. Stud. 18, 351–416.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2000. Great realizations. Risk 105–108.

Bandi, F.M., Russell, J.R., 2006. Separating microstructure noise from volatility. J. Financ. Econ. 79, 655–692.

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

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

Christensen, K., Oomen, R., Podolskij, M., 2014. Fact or friction: Jumps at ultra high frequency. J. Financ. Econ. 114, 576–599.

Delattre, S., Jacod, J., 1997. A central limit theorem for normalizd functions of the increments of a diffusion process in the presence of round-off errors. Bernoulli 3, 1–28.

Gloter, A., Jacod, J., 2001a. Diffusions with measurement errors. I–Local asymptotic normality. ESAIM Probab. Stat. 5, 225–242.

Gloter, A., Jacod, J., 2001b. Diffusions with measurement errors. II–Optimal estimators. ESAIM Probab. Stat. 5, 243–260.

Hall, P., Heyde, C.C., 1998. Martingale Limit Theory and its Application. Academic Press, Boston.

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econom. Statist. 24, 127–161.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the coninuous case: the pre-averaging approach. Stochastic Process. Appl. 119, 2249–2276.

Jacod, J., Li, Y., Zheng, X., 2017a. Statistical properties of microstructure noise. Econometrica 85, 1133–1174.

Jacod, J., Li, Y., Zheng, X., 2017b. Estimating the Integrated Volatility with Tick Observations. Working paper.

Jacod, J., Protter, P., 1998. Asymptotic error distributions for the euler method for stochastic differential equations. Ann. Probab. 26, 267–307.

Kolassa, J., McCullagh, P., 1990. Edgeworth series for lattice distributions. Ann. Statist. 18, 981–985.

Li, Y., Mykland, P.A., 2007. Are volatility estimators robust with respect to modeling assumptions? Bernoulli 13, 601–622.

Li, Y., Mykland, P.A., 2015. Rounding errors and volatility estimation. J. Financ. Econometrics 13, 478–504.

Li, Y., Xie, S., Zheng, X., 2016. Efficient estimation of integrated volatility incorporating trading information. J. Econometrics 195, 33–50.

Mykland, P.A., Zhang, L., 2006. ANOVA for diffusions and Itô processes. Ann. Statist. 34, 1931–1963.

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

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

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

Rosenbaum, M., 2009. Integrated volatility and round off error. Bernoulli 15, 687–720.

Strutz, T., 2010. Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond. Vieweg+Teubner Verlag, Wiesbaden.

Xiu, D., 2010. Quasi-maximum likellihood estimation of volatility with high frequency data. J. Econometrics 159, 235–250.

Zhang, L., 2001. From Martingales to ANOVA: Implied and Realized Volatility (Ph.D thesis), The University of Chicago, Department of Statistics.

Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale appraoch. Bernoulli 12, 1019–1043.

Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100, 1394–1411.

Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. J. Bus. Econom. Statist. 14, 45–52.