Journal of Econometrics 229 (2022) 422–451

Contents lists available at ScienceDirect

Journal of Econometrics

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

Volatility of volatility: Estimation and tests based on noisy high frequency data with jumps

Yingying Li [a,*,1], Guangying Liu [b,1], Zhiyuan Zhang [c,1]

[a] Department of ISOM and Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
[b] School of Statistics and Mathematics, Nanjing Audit University, No. 86 Yushan Western Road, Nanjing 211815, China
[c] School of Statistics and Management, Shanghai University of Finance and Economics, No. 777 Guoding Road, Shanghai 200433, China

ARTICLE INFO

Article history:
Received 1 January 2019
Received in revised form 31 December 2020
Accepted 4 February 2021
Available online 9 April 2021

JEL classification:
C14
C22
G12

Keywords:
Volatility of volatility
Central limit theorem
High frequency data
Microstructure noise
Semimartingale

ABSTRACT

We establish a feasible central limit theorem with convergence rate $n^{1/8}$ for the estimation of the integrated volatility of volatility (VoV) based on noisy high-frequency data with jumps. This is the first inference theory ever built for VoV estimation under such a general setup. The central limit theorem is applied to provide interval estimates of the VoV and conduct hypothesis tests. Furthermore, when one is interested in the null hypothesis that the VoV is zero, we show that a more powerful test can be established based on a VoV estimator with a convergence rate $n^{1/5}$ under the null. Empirical results on the S&P 500 and individual stocks show strong evidence of non-zero VoV.

1. Introduction

The increasing availability of high frequency data facilitates the understanding of characteristics of price dynamics such as volatility, jumps, and dependence structure of market microstructure noise. Taking volatility estimation as an example, in the past few decades, a large body of research has been devoted to answering the question of how to estimate volatilities more efficiently with high frequency data. In the case where the observed prices do not involve market microstructure noise, existing methods include the realized volatility (Andersen and Bollerslev, 1998; Barndorff-Nielsen and Shephard, 2002), realized bipower variation (Barndorff-Nielsen and Shephard, 2004), threshold realized volatility (Mancini, 2009), and threshold realized bipower volatility (Corsi et al., 2010). In the presence of microstructure noise, estimators include the two-scale realized volatility (Zhang et al., 2005), realized kernels (Barndorff-Nielsen et al., 2008), the pre-averaging method (Jacod et al., 2009; Podolskij and Vetter, 2009b), the quasi-maximum likelihood estimator (QMLE) (Xiu, 2010), the estimated-price realized volatility (Li et al., 2016), and the unified volatility estimator (Li et al., 2018). Additionally, Jacod et al. (2017) study the dependence structure of the market microstructure noise and propose in Jacod et al. (2019) a volatility estimator that applies directly to tick observations.

* Corresponding author.
E-mail addresses: yyli@ust.hk (Y. Li), liugying@nau.edu.cn (G. Liu), zhang.zhiyuan@mail.shufe.edu.cn (Z. Zhang).
1 Co-first authors.

https://doi.org/10.1016/j.jeconom.2021.02.007
0304-4076/

Volatility of volatility is a fundamental quantity that describes the dynamics of volatility processes. However, it is far less well-understood. Barndorff-Nielsen and Veraart (2009) propose an estimator of the *integrated volatility of volatility* (VoV hereafter) under a general nonparametric stochastic volatility model in the absence of microstructure noise. They first estimate the spot volatility based on the realized volatility over a local time window with window length shrinking to zero and then construct an estimator of the VoV which, up to a multiplicative correction, is a realized volatility of the estimated spot volatilities. Vetter (2015) proposes to correct the bias arising from the error in estimating the spot volatility, and constructs an estimator that achieves $n^{1/4}$ rate of convergence in the absence of noise. In the presence of microstructure noise, Sanfelici et al. (2015) develop an unbiased estimator of the VoV using Fourier analysis, albeit without establishing either the convergence rate or the central limit theorem (CLT). Under the assumption that the microstructure noise is a known parametric function of the limit order book information, Clinet and Potiron (2020) devise an estimator of the VoV which goes, in essence, to the noiseless case of Vetter (2015) and enjoys the same asymptotic properties as in Vetter (2015).

Besides the above early-stage explorations of VoV estimation, there is another strand of literature in understanding the dynamics of volatilities — various tests about the volatility processes, most of which are on parametric specifications. Aït-Sahalia (1996) develops tests for the spot interest rate model by comparing the density implied by the parametric specification against a nonparametric estimate of the density. Tests based on the (transition or marginal) density include those studied in Aït-Sahalia et al. (2009), Song (2011), Hong and Li (2005), Corradi and Swanson (2005), and Aït-Sahalia and Park (2012). These tests are based on low frequency observations, hence rely on long-span asymptotics. Under the high frequency data setting, Corradi and White (1999) adopt the variation method for specification tests of volatility function. Improved tests are proposed in Dette and von Lieres und Wilkau (2003), Dette et al. (2006), Dette and Podolskij (2008), Chen et al. (2019), and Bull (2017), among others. These articles are built on the assumption that the price is not contaminated by market microstructure noise. Vetter and Dette (2012), Papanicolaou and Giesecke (2016), and Christensen et al. (2018) extend these tests to the case where observations are contaminated by microstructure noise. Most of these tests assume under the null hypotheses that the (stochastic) volatility process is a diffusion process that is of a particular parametric form. Christensen et al. (2018) test the null hypothesis that the volatility process is a constant multiplied by a *deterministic* seasonal component, in which case the VoV is zero.

In this paper, we construct a nonparametric estimator of the VoV based on noisy high frequency data with jumps. We establish the convergence rate $n^{1/8}$ and build a feasible CLT. To the best of our knowledge, this is the first inference result for VoV under such a general setting.

In estimating the integrated volatility of the *price* process, several methods have been proposed, as discussed above. We provide insightful analogies between the estimation of integrated volatility and the estimation of VoV later in this paper to facilitate readers' understanding. However, we emphasize an essential difference between estimating the price process's integrated volatility and estimating the VoV. The difference lies in the fact that the (spot) volatility is not observable, not even in a contaminated form. We will first use the noisy data to build an estimator of the spot volatility and then use it to construct an estimator of the VoV. In broad terms, with noisy high frequency data observed at frequency n, we use observations in consecutive small blocks, each with length of time $O(n^{-1/4})$ to estimate spot volatilities, which can be viewed as the true spot volatilities contaminated by noise. A truncation method is used to deal with price jumps. Directly computing the realized variance based on these spot volatility estimates leads to an inconsistent estimate of the VoV. We resolve this inconsistency by correcting for the bias.

One can construct a test statistic based on the feasible CLT established above to test whether the VoV is at a particular level. Specifically, the test statistic is the VoV estimator standardized by its asymptotic mean and variance estimator under the null hypothesis. This test statistic converges (stably) in distribution to a standard normal with a rate of $n^{1/8}$ under the null hypothesis. Under the alternative hypothesis, the test statistic tends toward positive infinity at a rate of $n^{1/8}$.

Pushing one step further, we show that one could derive a more powerful test when one is interested in testing the null hypothesis that the VoV is zero. Under the usual continuous Itô semimartingale setup, if the volatility is truly free of a diffusion component, it will be smoother than the case where the volatility has a diffusion component. Intuitively, this extra smoothness facilitates the estimation of the spot volatility process, and consequently, should help detect whether the VoV is non-zero. We carry out this analysis rigorously and develop an estimator of VoV which, when scaled by $n^{1/5}$, converges (stably) in distribution to an asymptotic bias term plus a mixed normal variable under the null hypothesis. Based on this convergence result, we construct a new test statistic that diverges at a rate of $n^{1/5}$ under the alternative hypothesis while converging to a standard normal with a rate of $n^{1/5}$ under the null hypothesis.

The rest of this paper is organized as follows. We introduce the main results in three subsections of Section 2. In the first subsection, the spot volatility is estimated. In the second, the VoV is estimated, and in the third, the volatility process is tested to determine whether the VoV is zero. Simulation results are presented in Section 3. An empirical analysis is provided in Section 4, and we conclude in Section 5. All proofs and supplementary theoretical results are given in the Appendices.

2. Setup and main results

Suppose that the latent log price process is defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbb{P})$, and is a jump–diffusion process with stochastic volatility described by another diffusion process:

$$\begin{cases} dX_t = b_t\,dt + \sigma_t\,dW_t + dJ_t, \\ d\nu_t = \widetilde{b}_t\,dt + \widetilde{\sigma}_t\,d\widetilde{W}_t, \quad \nu_t = \sigma_t^2, \text{ for } t \in [0, T], \end{cases} \tag{1}$$

where b_t and $\widetilde{b}_t$ are drift processes; σ_t (or interchangeably, $\nu_t^{1/2}$) and $\widetilde{\sigma}_t$ are volatility processes; W_t and $\widetilde{W}_t$ are standard Brownian motions which may be correlated with each other; and J_t is a pure jump process. For any t, σ_t represents the spot volatility, and $\widetilde{\sigma}_t$ represents the spot volatility of volatility. We emphasize that the assumption that the spot volatility is a diffusion process is widely adopted in the continuous-time volatility literature, see, e.g., Heston (1993), Barndorff-Nielsen and Veraart (2009), Aït-Sahalia et al. (2013), Vetter (2015) and Mykland and Zhang (2017).

Observations are contaminated versions of the latent log price process at discrete times:

$$Y_{t_i} = X_{t_i} + \varepsilon_i, \tag{2}$$

where t_i's are observation times in a fixed time interval $[0, T]$, and ε_i represents the market microstructure noise that is used to capture the bid–ask spread, discreteness of price changes etc. In this paper, we only consider the case in which the observation times are equally spaced, i.e., $t_i = i/n$ for $i = 0, 1, \ldots [nT]$, where $[x]$ denotes the integer part of x.

The assumptions about the observed log price process (2) are more specifically summarized as follows:

Assumption 1. The observed log price process and the latent log price process are given by (2) and (1), respectively.

(i) The drift processes $b, \widetilde{b}$ and the volatility processes $\sigma, \widetilde{\sigma}$ are all adapted, locally bounded, and càdlàg. $\sigma_t > 0$ and $\widetilde{\sigma}_t \geq 0$ for all $t \geq 0$. W and $\widetilde{W}$ are two standard Brownian motions that may be correlated.

(ii) The jump process J follows

$$J_t = \delta * p_t,$$

where p is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{E}$ with $(\mathbb{E}, \mathcal{E})$ an auxiliary Polish space, and with compensator $dt \otimes q(dz)$; and δ is a predictable function on $\Omega \times \mathbb{R}_+ \times \mathbb{E}$. For some $0 < \beta < 1$, $|\delta(\omega, t, z)|^\beta \leq \overline{\delta}_n(z)$ whenever $t \leq \tau_n(\omega)$, where τ_n is a sequence of stopping times increasing to ∞, and $\overline{\delta}_n(z)$ is a sequence of deterministic functions satisfying $\int \overline{\delta}_n(z)q(dz) < \infty$.

(iii) The noise terms ε_i, $i = 0, 1, \ldots [nT]$ are independently and identically distributed, and are independent of the latent log price and volatility processes, with mean zero, variance σ_ϵ^2, and finite moments of all orders.

The above assumptions with respect to the nonparametric stochastic volatility model are classical and often used in analyzing high frequency data; see, for example, Vetter (2015), Wang and Mykland (2014), Aït-Sahalia and Jacod (2014), and Xiu (2010). The assumptions about J suggest that it is a bounded variation process.

2.1. Estimating spot volatility

VoV is an important variable in finance. First, it reflects how stable (or volatile) the volatility is. Second, it is an important variable in pricing volatility derivatives; see, for example, Cont and Kokholm (2013). Third, it is a basic variable for studying other important quantities such as the leverage effect, which refers to the generally negative correlation between asset returns and changes in volatility. Similar to the integrated volatility, which gives an overall measure of how volatile the price process is, the VoV gives an overall measure of how volatile the volatility process is. One of our aims is to construct an estimator of the VoV, i.e., $\int_0^T \widetilde{\sigma}_t^2 dt$, based on noisy high frequency data with jumps.

The VoV is indeed the quadratic variation of the volatility process, i.e., $\int_0^T \widetilde{\sigma}_s^2 ds = \langle \nu, \nu \rangle_T$. Furthermore, the classical theory using high frequency data shows that methods such as realized volatility and the pre-averaging method can estimate the VoV if the process ν is observable (without or with noise). However, the spot volatilities ν_{t_i}'s are not observable. A natural path is to first build an estimator $\widehat{\nu}_{t_i}$ for the spot volatility, and then to construct an estimator $\widehat{\langle \nu, \nu \rangle}_t$ based on the estimated values $\widehat{\nu}_{t_i}$. Foster and Nelson (1996) show that the spot volatility could be estimated from high frequency data using rolling and block sampling filters. Related methods include those of Fan and Wang (2008) and Liu et al. (2017) in the noiseless high frequency data setting, and the methods of Zu and Boswijk (2014), Mancini et al. (2015), and Mancino et al. (2017) for noisy high frequency data.

To construct an estimator of the spot volatility based on log prices contaminated by market microstructure noise, we recall the pre-averaging method, which involves a kernel function and a moving window. Let $g(x), x \in [0, 1]$ be the kernel function and k_n be the number of observations in the moving window. The kernel function g is chosen to satisfy the following properties:

$$\begin{cases} g \text{ is continuous, piecewise } C^1 \text{ with a piecewise Lipschitz derivative } g', \\ g(0) = g(1) = 0, \quad \displaystyle\int_0^1 g(s)^2 ds > 0. \end{cases}$$

We define some quantities and functions on $\mathbb{R}^+$ associated with the weight function g:

$$\begin{cases} s \in [0, 1] \mapsto \phi_1(s) = \displaystyle\int_s^1 g'(u)g'(u-s)du, \quad \phi_2(s) = \int_s^1 g(u)g(u-s)du, \\ s > 1 \mapsto \phi_1(s) = \phi_2(s) = 0, \\ i, j = 1, 2 \Rightarrow \Phi_{ij} = \displaystyle\int_0^1 \phi_i(s)\phi_j(s)ds, \quad \psi_i = \phi_i(0). \end{cases}$$

For any process V, set

$$\overline{V}_i^n = \sum_{j=1}^{k_n-1} g_j^n \Delta_{i+j}^n V, \quad g_j^n = g(j/k_n), \quad \Delta_i^n V = V_{i/n} - V_{(i-1)/n}.$$

Note that $\overline{V}_i^n$ is a local weighted average of increments of V. In our simulation and empirical studies, we adopt the triangular kernel $g(x) = x \wedge (1-x)$ which leads to

$$\overline{V}_i^n = \frac{1}{2}\left(\frac{1}{k_n'}\sum_{j=0}^{k_n'-1} V_{i+j+k_n'}^n - \frac{1}{k_n'}\sum_{j=0}^{k_n'-1} V_{i+j}^n\right),$$

for $k_n = 2k_n'$ where k_n' is an integer, and

$$\psi_1 = 1, \quad \psi_2 = \frac{1}{12}, \quad \Phi_{11} = \frac{1}{6}, \quad \Phi_{12} = \frac{1}{96}, \quad \Phi_{22} = \frac{151}{80640}.$$

However, throughout the derivation of theoretical results, g is assumed to be a general function that satisfies the conditions mentioned above.

Jing et al. (2014) propose the threshold pre-averaging estimator of the integrated volatility. Let l_n be the number of price increments (with time step $1/n$) in a local window that is wider than the pre-averaging moving window (i.e., $l_n > k_n$). The threshold pre-averaging estimator over the time interval $[i/n, (i+l_n)/n]$ is given by

$$IV_{i,l_n}^{PA} = \frac{1}{\psi_2 k_n}\sum_{j=0}^{l_n-k_n+1}(\overline{Y}_{i+j}^n)^2 1_{|\overline{Y}_{i+j}^n|<u_n} - \frac{\psi_1}{2\psi_2 k_n^2}\sum_{j=0}^{l_n}(\Delta_{i+j}^n Y)^2, \tag{3}$$

where u_n is a threshold value to remove the impact of jumps. They show that $IV_{0,[nt]}^{PA}$ is a consistent estimator of the integrated variance $\int_0^t \nu_s ds$. A natural estimator for the spot variance ν_t is then given by

$$\widehat{\nu}_t = \frac{1}{l_n/n} IV_{i,l_n}^{PA}, \quad \text{for } t \in ((i-1)/n, i/n]. \tag{4}$$

Note that the observations involved to build $\widehat{\nu}_t$ are within the time interval $[i/n, (i+l_n)/n]$.

In the sequel, we set

$$l_n = [c_1 n^{\theta_1}], \quad u_n = c_2 n^{-\theta_2}, \quad k_n = [c_3 n^{1/2}], \quad \theta_1 \geq 1/2,\ \theta_2 < 1/4, \quad c_i > 0, \quad i = 1, 2, 3, \tag{5}$$

where (c_i) are fixed constants. Proposition 1 gives the asymptotics of $\widehat{\nu}_t$ which we derive based on the results on page 296 of Aït-Sahalia and Jacod (2014) where the latent price process is assumed to be continuous. Before presenting our result, we define two processes Z and Z' on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ of the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Conditional on $\mathcal{F}$, these processes are independent Gaussian white noise processes with the following conditional variances:

$$\widetilde{E}(Z_t^2|\mathcal{F}) = \gamma_t^2, \quad \widetilde{E}(Z_t'^2|\mathcal{F}) = \widetilde{\sigma}_t^2,$$

where

$$\gamma_t^2 = \frac{4}{\psi_2^2}\left[\Phi_{22}c_3\sigma_t^4 + \frac{2\Phi_{12}}{c_3}\sigma_\epsilon^2\sigma_t^2 + \frac{\Phi_{11}}{c_3^3}\sigma_\epsilon^4\right]. \tag{6}$$

Note that $\int_0^t \gamma_s^2 ds$ is the asymptotic variance of the pre-averaging estimator (Jacod et al., 2009).

We also need the concept of *stable convergence* in law. Suppose that a sequence of variables (Y_n) is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a sub-sigma-field $\mathcal{G} \subseteq \mathcal{F}$, and that the limiting variable Y is defined on an extended probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$. We say that Y_n converges $\mathcal{G}$-stably in law to Y if, for all bounded $\mathcal{G}$-measurable variables Z and all bounded functions g, $Eg(Y_n)Z \to \widetilde{E}g(Y)Z$. If $\mathcal{G} = \mathcal{F}$, we simply say that Y_n converges stably in law to Y and write $Y_n \xrightarrow{L-s} Y$. For more details, see, for example, Jacod and Shiryaev (2003).

Proposition 1. *Suppose that Assumption 1 holds.*

(i) If $1/2 < \theta_1 < 3/4$ *and* $(2\theta_1 - 1)/(4(2-\beta)) < \theta_2 < 1/4$*, then for any fixed* t

$$l_n^{1/2} n^{-1/4}(\widehat{\nu}_t - \nu_t) \xrightarrow{L-s} Z_t.$$

(ii) If $\theta_1 = 3/4$ *and* $1/(8(2-\beta)) < \theta_2 < 1/4$*, then for any fixed* t

$$l_n^{1/2} n^{-1/4}(\widehat{\nu}_t - \nu_t) \xrightarrow{L-s} Z_t + c_1 Z_t'. \tag{7}$$

(iii) If $3/4 < \theta_1 < 1$ and $(1-\theta_1)/(2(2-\beta)) < \theta_2 < 1/4$, then for any fixed t

$$\sqrt{\frac{n}{l_n}}(\widehat{\nu}_t - \nu_t) \xrightarrow{L-s} Z'_t.$$

Remark 1. The fastest convergence rate corresponds to the second case when $\theta_1 = 3/4$, and the rate is $n^{1/8}$. The same rate has been obtained by Munk and Schmidt-Hieber (2010) when volatility processes meet suitable Hölder-smoothness and *uniform boundedness* conditions. Notice that Munk and Schmidt-Hieber (2010) derive minimax lower bounds for the estimation of the instantaneous volatility without establishing central limit theorems. Moreover, under a general setup which considers both price and volatility jumps, an almost optimal rate $n^{1/8-\iota}$ (for any $\iota > 0$) was obtained by Bibinger and Winkelmann (2018) when volatility consists of both an Itô semimartingale and a component that has Hölder continuous paths with Hölder exponent less than or equal to 1/2. If $\widetilde{\sigma}_s \equiv 0$ and ν_t is a bounded variation process, then it is not difficult to prove that for any fixed t

$$l_n^{1/2} n^{-1/4}(\widehat{\nu}_t - \nu_t) \xrightarrow{L-s} Z_t + c_1^{3/2}\widetilde{b}_t/2$$

when $\theta_1 = 5/6$. In particular, the convergence rate is $n^{1/6}$.

2.2. Estimating the VoV

We now construct an estimator of the VoV. Set $\widetilde{Z}_t = Z_t + c_1 Z'_t$, and by the convergence (7), we have

$$\widehat{\nu}_t = \nu_t + \widetilde{\epsilon}_t, \quad \text{where} \quad n^{1/8}\widetilde{\epsilon}_t \xrightarrow{L-s} \frac{1}{\sqrt{c_1}}\widetilde{Z}_t, \quad \text{for} \quad t = il_n/n, i = 1, \ldots, [nT/l_n].$$

In other words, $\widehat{\nu}$ can be regarded as the underlying signal process ν contaminated by noise $\widetilde{\epsilon}_t$. Our goal is to estimate $\langle \nu, \nu \rangle_t$. We therefore arrive at a situation which is reminiscent of the classical one of estimating the integrated volatility based on noisy data. But ν_t and $\widetilde{\epsilon}_t$ are dependent, making the estimation of VoV challenging and our analysis innovative. Notice that the inference theory for estimating the VoV in the presence of microstructure noise and price jumps has not been established in the prior literature. Before presenting our VoV estimator, we shall first investigate the properties of the realized volatility of the estimated spot volatility process.

For an integer $m \geq 1$, define

$$V(\nu; \theta_1, m)_t^n = \sum_{i=0}^{[t/(l_n/n)]-m-1} (\widehat{\nu}_{(i+m)l_n/n} - \widehat{\nu}_{il_n/n})^2. \tag{8}$$

Recall (3)–(5) for the definitions of $\widehat{\nu}$, l_n and the jump truncation threshold u_n. We have the following proposition.

Proposition 2. *Suppose that Assumption 1 holds, $\theta_1 = 3/4$ and $1/(8(2-\beta)) < \theta_2 < 1/4$. Then,*

$$V(\nu; \theta_1, m)_t^n \xrightarrow{P} (m - 1/3)\int_0^t \widetilde{\sigma}_s^2 ds + \frac{2}{c_1^2}\int_0^t \gamma_s^2 ds. \tag{9}$$

Remark 2. In the case where prices contain no microstructure noise and jumps, the classical realized volatility theory states that $\sum_{i=1}^{[nt]}(X_{i/n} - X_{(i-1)/n})^2 \xrightarrow{P} \int_0^t \sigma_s^2 ds$. Comparing this classical limiting result with Proposition 2, one notices that there are two main differences in their limiting variables. The first difference is that there is a multiplicator $m - 1/3$ in the term involving the integrated volatility (of volatility). In fact, if $\widehat{\nu}_{(i+m)l_n/n}$ and $\widehat{\nu}_{il_n/n}$ in $V(\nu; \theta_1, m)_t^n$ are replaced by $\int_{(i+m)l_n/n}^{(i+m+1)l_n/n} \nu_s ds/(l_n/n)$ and $\int_{il_n/n}^{(i+1)l_n/n} \nu_s ds/(l_n/n)$, respectively, then the multiplicator $m - 1/3$ still exists in the limiting variable, whereas the term $2\int_0^t \gamma_s^2 ds/c_1^2$ disappears. The multiplicator is due to the use of a local average $\int_t^{t+l_n/n} \nu_s ds/(l_n/n)$ to estimate ν_t. The second difference is the additional term $2\int_0^t \gamma_s^2 ds/c_1^2$ in Proposition 2 compared with the classical result. The term $2\int_0^t \gamma_s^2 ds/c_1^2$ comes from the error induced by estimating $\int_{il_n/n}^{(i+1)l_n/n} \nu_s ds/(l_n/n)$ via $\widehat{\nu}_{il_n/n}$.

From Proposition 2, one can construct a consistent estimator for $\langle \nu, \nu \rangle_t = \int_0^t \widetilde{\sigma}_s^2 ds$ provided that the term $\int_0^t \gamma_s^2 ds$ can be estimated. To build a consistent estimator of $\int_0^t \gamma_s^2 ds$, let

$$V(Y, p)_t^n = n^{p/4-1}\sum_{i=0}^{[nt]-k_n+1} |\overline{Y}_i^n|^p 1_{|\overline{Y}_i^n| < u_n} \quad \text{and} \widehat{\sigma_\epsilon^2} = \frac{1}{2nt}\sum_{i=0}^{[nt]-1}(\Delta_i^n Y)^2.$$

By Lemma A.1 of Jing et al. (2014), suppose that $\max\{(p-2)/(4(p-\beta)), 0\} < \theta_2 < 1/4$, then

$$V(Y, p)_t^n \xrightarrow{P} \mu_p \int_0^t \left(c_3\psi_2\sigma_s^2 + \frac{1}{c_3}\psi_1\sigma_\epsilon^2\right)^{\frac{p}{2}} ds, \tag{10}$$

where μ_p is the pth order absolute moment of a standard normal random variable. Additionally, $\widehat{\sigma_\epsilon^2} \xrightarrow{P} \sigma_\epsilon^2$. Based on these results, we build a consistent estimator of $\int_0^t \gamma_s^2 ds$ as

$$\widehat{\int_0^t \gamma_s^2 ds} = \frac{4\Phi_{22}}{3c_3\psi_2^4} V(Y,4)_t^n + \frac{8}{c_3^2}\left(\frac{\Phi_{12}}{\psi_2^3} - \frac{\Phi_{22}\psi_1}{\psi_2^4}\right) V(Y,2)_t^n \widehat{\sigma_\epsilon^2} + \frac{4t}{c_3^3}\left(\frac{\Phi_{11}}{\psi_2^2} - 2\frac{\Phi_{12}\psi_1}{\psi_2^3} + \frac{\Phi_{22}\psi_1^2}{\psi_2^4}\right)(\widehat{\sigma_\epsilon^2})^2.$$

Finally, we construct our estimator of $\langle \nu, \nu \rangle_t$ as

$$\widehat{\langle \nu, \nu \rangle}_t = \frac{3}{2} V(\nu; 3/4, 1)_t^n - \frac{3}{c_1^2} \widehat{\int_0^t \gamma_s^2 ds}. \tag{11}$$

Remark 3. From the convergence result (9) in Proposition 2, one can see that when $m \to \infty$, $V(\nu; 3/4, m)_t^n/(m - 1/3)$, which is guaranteed to be positive, is a consistent estimator of the VoV. The bias correction in (11) is not necessary if m can be chosen sufficiently large. We keep the theory in its general form by allowing m to take different values because this has further methodological implications. Nonetheless, because the estimator with $m = 1$ gives the best finite-sample performance in numerical studies, we present the VoV estimator in the form of $m = 1$ in the main text and implement it in subsequent simulation and empirical studies.

Remark 4. In the presence of (finite activity) volatility jumps, similarly to how we obtain an integrated volatility estimator (3) robust to price jumps, one could impose a further truncation at the spot volatility level in (8). Specifically on the summands (i.e., the squared estimated spot volatility changes) to make the VoV estimator robust to volatility jumps.

To establish the CLT of $\widehat{\langle \nu, \nu \rangle}_t$, we must make some additional assumptions. The following assumption is used in Vetter (2015) to build the distribution theory for the estimator of the VoV based on noiseless data.

Assumption 2. Suppose that Assumption 1 holds and that the diffusion process $\widetilde{\sigma}$ has the following form:

$$d\widetilde{\sigma}_t = \overline{b}_t dt + \overline{\sigma}_t d\overline{W}_t,$$

where $\overline{b}$ and $\overline{\sigma}$ are two adapted, locally bounded and càdlàg processes, and $\overline{W}$ is a standard Brownian motion that may be correlated with W and $\widetilde{W}$.

Theorem 1. *Suppose that Assumption 1 holds and $1/(8(2-\beta)) < \theta_2 < 1/4$. Then,*

$$\widehat{\langle \nu, \nu \rangle}_t \xrightarrow{P} \langle \nu, \nu \rangle_t.$$

Furthermore, suppose that Assumption 2 holds and $3/(16(2-\beta)) < \theta_2 < 1/4$. Then, for any fixed t,

$$n^{1/8}\left(\widehat{\langle \nu, \nu \rangle}_t - \langle \nu, \nu \rangle_t\right) \xrightarrow{L-s} \int_0^t \Sigma_s dB_s, \tag{12}$$

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

$$\Sigma_t^2 = \frac{9c_1}{4}\widetilde{\sigma}_t^4 + \frac{9}{c_1}\widetilde{\sigma}_t^2\gamma_t^2 + \frac{27}{c_1^3}\gamma_t^4 \tag{13}$$

with γ_t^2 being defined in (6).

Remark 5. Similar to the discussions in Remark 2 of Jacod et al. (2009), we may obtain an asymptotic variance that is close to the parametric efficiency bound of Gloter and Jacod (2001). However, clear discussions may only be possible when σ_t is constant. This means the VoV is equal to zero, which is not an interesting case for this study. For this reason, we do not pursue a further theoretical analysis of the optimal choices of the constants c_3 and c_1. In the empirical studies, we provide a data-driven procedure to choose the constant c_1.

Remark 6. When the latent process X is continuous, Vetter (2015) constructs an estimator for the VoV in the noiseless case, where the consistency and asymptotic properties of the estimator have been established. In a noisy setting where the microstructure noise is a known parametric function of trading information, Clinet and Potiron (2020) show that the same asymptotic properties in Vetter (2015) can be achieved in estimating the VoV. Sanfelici et al. (2015) use the Fourier method to estimate the VoV for both noisy and noiseless cases. They demonstrate the consistency of the estimator in the noiseless case and the unbiasedness of the estimator in the noisy case, albeit without establishing either the convergence rate or the associated CLT. Furthermore, to the best of our knowledge, estimation of VoV has not been studied in the case where X contains jumps.

Remark 7. In the noiseless case, the VoV estimator of Vetter (2015) achieves a rate of convergence $n^{1/4}$. In a noisy setting without price jumps, Wang and Mykland (2014) study the estimation of the leverage effect, which is defined as the covariation between the volatility process and the latent return process $\langle X, \nu\rangle_t$, and develop a consistent estimator with a rate of convergence $n^{1/8}$. Mathematically, $\langle X, \nu\rangle_t$ is a third moment quantity, while our target $\langle \nu, \nu\rangle_t$ is a fourth moment quantity whose estimation is more involved.

The limiting distribution of Theorem 1 is a mixed normal. To obtain a feasible CLT, we need to build a consistent estimator of the conditional variance $\int_0^t \Sigma_s^2 ds$. For a positive integer $m \geq 1$, let

$$PV(\nu, m)_t^n = n^{1/4} \sum_{i=1}^{[nt/l_n]-(m+1)} (\widehat{\nu}_{(i+m)l_n/n} - \widehat{\nu}_{il_n/n})^4,$$

$$BV(\nu)_t^n = n^{1/4} \sum_{i=1}^{[nt/l_n]-3} (\widehat{\nu}_{(i+1)l_n/n} - \widehat{\nu}_{il_n/n})^2 (\widehat{\nu}_{(i+2)l_n/n} - \widehat{\nu}_{(i+1)l_n/n})^2.$$

By the same methods as in the proof of Proposition 2, suppose that Assumption 1 holds, $l_n = [c_1 n^{3/4}]$ and $1/(8(2-\beta)) < \theta_2 < 1/4$, then we have the following convergence in probability:

$$PV(\nu, m)_t^n \xrightarrow{P} \frac{(3m-1)^2 c_1}{3} \int_0^t \widetilde{\sigma}_s^4 ds + \frac{12m-4}{c_1} \int_0^t \widetilde{\sigma}_s^2 \gamma_s^2 ds + \frac{12}{c_1^3} \int_0^t \gamma_s^4 ds, \tag{14}$$

$$BV(\nu)_t^n \xrightarrow{P} \frac{c_1}{2} \int_0^t \widetilde{\sigma}_s^4 ds + \frac{2}{c_1} \int_0^t \widetilde{\sigma}_s^2 \gamma_s^2 ds + \frac{6}{c_1^3} \int_0^t \gamma_s^4 ds. \tag{15}$$

Thus, we have the following feasible CLT.

Theorem 2. *Suppose that Assumption 2 holds and $3/(16(2-\beta)) < \theta_2 < 1/4$. Then,*

$$n^{\frac{1}{8}} \frac{\widehat{\langle \nu, \nu\rangle}_t - \langle \nu, \nu\rangle_t}{\sqrt{\Xi_t^n}} \xrightarrow{L-s} N(0,1), \tag{16}$$

where

$$\Xi_t^n = 9BV(\nu)_t^n/2. \tag{17}$$

Remark 8. From (14), one can construct another consistent estimator of $\int_0^t \Sigma_s^2 ds$:

$$7PV(\nu, 1)_t^n - \frac{29}{4} PV(\nu, 2)_t^n + \frac{5}{2} PV(\nu, 3)_t^n \xrightarrow{P} \int_0^t \Sigma_s^2 ds. \tag{18}$$

Our simulation results show that the performance of (18) is not as good as (17), and sometimes the estimator in (18) yields negative values. Mykland and Zhang (2017) propose a two-scale method to estimate the asymptotic variance for a broad class of integrated parameters. However, applying their asymptotic variance estimator here may also produce negative values.

2.3. Testing whether a pure drift process gives the volatility process

Our estimation theory established in the previous section can be readily applied to either construct confidence intervals or build a test against a null hypothesis that the VoV is at a particular level. In this subsection, we show that what we can obtain is beyond a straightforward application of our feasible CLT if one is interested in testing against the null hypothesis that VoV is zero. On the one hand, the VoV is zero in popular price models proposed by, e.g., Black and Scholes (1973) and Vasicek (1977). On the other hand, in the literature of testing parametric specification of volatility as discussed in the Introduction section, volatility processes under the null hypotheses except that of Christensen et al. (2018) assume particular parametric forms with either non-zero diffusion components or zero VoV. The null Christensen et al. (2018) test is given by a constant multiplied by a (differentiable) deterministic seasonal component; thus, the VoV is zero. Therefore, an interesting question arises naturally: would drift processes alone be sufficient in modeling volatility processes? In other words, the test problem in our context is: does $\int_0^t \widetilde{\sigma}_s^2 ds = 0$ or not? We solve this problem from the perspective of VoV. The hypotheses to be tested are formally stated as follows:

$$H_0: P\left(\{\widetilde{\sigma}_s^2 = 0 \text{ for all } s \in [0,t]\}\right) = 1 \quad \text{vs.} \quad H_1: P\left(\left\{\int_0^t \widetilde{\sigma}_s^2 ds > 0\right\}\right) = 1.$$

Here, H_0 means the volatility is a pure drift process, while H_1 means the volatility consists of a non-zero diffusion component under the setup of this paper.

Under the null hypothesis, Theorem 1 implies that $\widehat{\langle \nu, \nu \rangle}_t \xrightarrow{P} 0$ and $\Sigma_t^2 = 27\gamma_t^4/c_1^3$. From (10), we can then build a consistent estimator of $\int_0^t \gamma_s^4 ds$:

$$\begin{aligned}
&\widehat{\int_0^t \gamma_s^4 ds} \\
:=& \frac{16}{\psi_2^4}\Bigg[\frac{\Phi_{22}^2}{105c_3^2\psi_2^4}V(Y,8)_t^n + \frac{4}{15c_3^3}\left(-\frac{\Phi_{22}^2\psi_1}{\psi_2^4} + \frac{\Phi_{11}\Phi_{12}}{\psi_2^3}\right)V(Y,6)_t^n\widehat{\sigma_\epsilon^2} \\
&+ \frac{2}{c_3^4}\left(\frac{\Phi_{22}^2\psi_1^2}{\psi_2^4} - 2\frac{\Phi_{11}\Phi_{22}\psi_1}{\psi_2^3} + \frac{1}{3}\frac{2\Phi_{12}^2 + \Phi_{11}\Phi_{22}}{\psi_2^2}\right)V(Y,4)_t^n(\widehat{\sigma_\epsilon^2})^2 \\
&+ \frac{4}{c_3^5}\left(-\frac{\Phi_{22}^2\psi_1^3}{\psi_2^4} + 3\frac{\Phi_{11}\Phi_{22}\psi_1^2}{\psi_2^3} - \frac{(2\Phi_{12}^2 + \Phi_{11}\Phi_{22})\psi_1}{\psi_2^2} + \frac{\Phi_{11}\Phi_{12}}{\psi_2}\right)V(Y,2)_t^n(\widehat{\sigma_\epsilon^2})^3 \\
&+ \frac{(\widehat{\sigma_\epsilon^2})^4 t}{c_3^6}\left(\frac{\Phi_{22}^2\psi_1^4}{\psi_2^4} - 4\frac{\Phi_{11}\Phi_{22}\psi_1^3}{\psi_2^3} + 2\frac{(2\Phi_{12}^2 + \Phi_{11}\Phi_{22})\psi_1^2}{\psi_2^2} - 4\frac{\Phi_{11}\Phi_{12}\psi_1}{\psi_2} + \Phi_{11}^2\right)\Bigg].
\end{aligned} \tag{19}$$

The consistency of (19) holds under both the null and alternative hypotheses. Then, based on Theorem 1, we obtain a preliminary test statistic

$$\breve{T}_n := n^{\frac{1}{8}} \frac{\widehat{\langle \nu, \nu \rangle}_t}{\sqrt{27\widehat{\int_0^t \gamma_s^4 ds}/c_1^3}} \tag{20}$$

and have that $\breve{T}_n \xrightarrow{L} N(0,1)$ under the null hypothesis that the volatility process is a bounded variation process and $\langle \nu, \nu \rangle_t = 0$. Theorem 1 also implies that the above test statistic goes to $+\infty$ at the rate of $n^{1/8}$ under the alternative hypothesis.

Can the test be improved in terms of convergence rate? An important observation is that, under the null hypothesis, the diffusion term in the volatility process disappears and the spot volatility becomes a bounded variation process. Intuitively, if the volatility is indeed a bounded variation process, then it will be smoother than the usual setting when the volatility is an Itô diffusion process. This extra smoothness should facilitate the estimation of the spot volatility process. We next make this intuition rigorous. To this end, we introduce the following notation

$$\widehat{\langle \nu, \nu \rangle}_t^{(\theta_1)} := \frac{3}{2}\left(V(\nu; \theta_1, 1)_t^n - \frac{2n^{\frac{3}{2}-2\theta_1}}{c_1^2}\widehat{\int_0^t \gamma_s^2 ds}\right). \tag{21}$$

As can be seen from the convergence result (A.28) in Appendix A, $\widehat{\langle \nu, \nu \rangle}_t^{(\theta_1)}$ is a consistent estimator of $\int_0^t \widetilde{\sigma}_s^2 ds$ as long as the conditions that $1/2 < \theta_1 < 1$ and $(1-\theta_1)/(4-2\beta) < \theta_2 < 1/4$ hold. Obviously, when $\theta_1 = 3/4$, $\widehat{\langle \nu, \nu \rangle}_t^{(3/4)}$ reduces to our proposed VoV estimator $\widehat{\langle \nu, \nu \rangle}_t$. Moreover, from the discussion of Remark 12 in Appendix A, it is easily seen that, when $\int_0^t \widetilde{\sigma}_s^2 ds = 0$, the highest convergence rate in estimating the VoV (which is zero in this case) is achieved by $\widehat{\langle \nu, \nu \rangle}_t^{(\theta_1)}$ with $\theta_1 = 4/5$, whose limiting results are summarized into the following proposition.

Proposition 3. *Suppose that Assumption 1 holds, $\widetilde{\sigma} \equiv 0$, and $1/(5(2-\beta)) < \theta_2 < 1/4$. Then, for any fixed $t > 0$,*

$$n^{\frac{1}{5}}\widehat{\langle \nu, \nu \rangle}_t^{(4/5)} \xrightarrow{L-s} \frac{3c_1}{2}\int_0^t \widetilde{b}_s^2 ds + \sqrt{\frac{27}{c_1^3}}\int_0^t \gamma_s^2 dB_s,$$

where $\widehat{\langle \nu, \nu \rangle}_t^{(4/5)}$ is given by (21) with $\theta_1 = 4/5$.

Based on Proposition 3, we define the following infeasible test statistic

$$\widetilde{T}_n =: \frac{n^{1/5}\widehat{\langle \nu, \nu \rangle}_t^{(4/5)} - \frac{3c_1}{2}\int_0^t \widetilde{b}_s^2 ds}{\sqrt{\frac{27}{c_1^3}\int_0^t \gamma_s^4 ds}} \tag{22}$$

and have that $\widetilde{T}_n \xrightarrow{L} N(0,1)$ under the null hypothesis. Under the alternative hypothesis that $\int_0^t \widetilde{\sigma}_s^2 ds > 0$, the convergence (A.28) in Appendix A implies that

$$\widetilde{T}_n = \left(\frac{\int_0^t \widetilde{\sigma}_s^2 ds}{\sqrt{\frac{27}{c_1^3} \int_0^t \gamma_s^4 ds}} + o_P(1) \right) \times n^{1/5} + O_P(1),$$

which diverges at the rate of $n^{1/5}$. This rate is faster than the $n^{1/8}$ of the preliminary test statistic $\breve{T}_n$ in (20). To make $\widetilde{T}_n$ feasible, we need to consistently estimate the term $27/c_1^3 \int_0^t \gamma_s^4 ds$ in the denominator and the term $3c_1 \int_0^t \widetilde{b}_s^2 ds/2$ in the numerator. We treat them one by one as follows.

First, we have seen that $\widehat{\int_0^t \gamma_s^4 ds}$ defined in (19) consistently estimates $\int_0^t \gamma_s^4 ds$ under both the null and alternative hypotheses. In fact, under the null hypothesis, by (15), $9BV(\nu)_t^n/2$ (with $\theta_1 = 3/4$) is a consistent estimator of $27 \int_0^t \gamma_s^4 ds/c_1^3$, whereas it converges to a limit greater than $27 \int_0^t \gamma_s^4 ds/c_1^3$ under the alternative hypothesis. Therefore,

$$\widehat{\Upsilon}_t := \min \left\{ \frac{27}{c_1^3} \widehat{\int_0^t \gamma_s^4 ds}, \frac{9}{2} BV(\nu)_t^n \right\} \tag{23}$$

with $\theta_1 = 3/4$ consistently estimates $27/c_1^3 \int_0^t \gamma_s^4 ds$ under both the null and alternative hypotheses. Our unreported simulation results show that, in finite samples, the test based on (23) outperforms that based solely on $27/c_1^3 \widehat{\int_0^t \gamma_s^4 ds}$ in terms of powers while the two tests have comparable sizes.

Second, under the null hypothesis, we can construct a (nonnegative) consistent estimator of $\int_0^t \widetilde{b}_s^2 ds$ as

$$\widehat{\mathcal{B}}_t := \widehat{\int_0^t \widetilde{b}_s^2 ds} \times 1_{\widehat{\int_0^t \widetilde{b}_s^2 ds} \geq 0}, \tag{24}$$

where $\widehat{\int_0^t \widetilde{b}_s^2 ds}$ is defined as follows

$$\widehat{\int_0^t \widetilde{b}_s^2 ds} := \frac{n^{1/6}}{3c_1} \left(V(\nu; 5/6, 2)_t^n - V(\nu; 5/6, 1)_t^n \right),$$

which, from result (A.29) in Appendix A, is a consistent estimator of $\int_0^t \widetilde{b}_s^2 ds$, provided that $\widetilde{\sigma} \equiv 0$. We adopt $\widehat{\mathcal{B}}_t$ because $\widehat{\int_0^t \widetilde{b}_s^2 ds}$ may produce negative values in finite samples. Under the alternative hypothesis, the convergence result (A.28) in Appendix A implies that

$$V(\nu; 5/6, 2)_t^n - V(\nu; 5/6, 1)_t^n \xrightarrow{P} \int_0^t \widetilde{\sigma}_s^2 ds$$

and hence $\widehat{\int_0^t \widetilde{b}_s^2 ds}$ and $\widehat{\mathcal{B}}_t$ are of order $O_P(n^{1/6})$.

We have thus obtained the following feasible test statistic:

$$T_n := \frac{n^{1/5} \widehat{\langle \nu, \nu \rangle}_t^{(4/5)} - \frac{3c_1}{2} \widehat{\mathcal{B}}_t}{\sqrt{\widehat{\Upsilon}_t}}. \tag{25}$$

The asymptotic property of T_n under the null hypothesis follows easily from that of $\widetilde{T}_n$. We now briefly explain the asymptotic behavior of T_n under the alternative hypothesis as this may not be directly clear. First, because

$$\widehat{\langle \nu, \nu \rangle}_t^{(4/5)} \xrightarrow{P} \int_0^t \widetilde{\sigma}_s^2 ds$$

as indicated by the discussion immediately following (21), we have

$$n^{1/5} \widehat{\langle \nu, \nu \rangle}_t^{(4/5)} = \left(\int_0^t \widetilde{\sigma}_s^2 ds + o_P(1) \right) \times n^{1/5}$$

under the alternative hypothesis. Second, the discussion following (24) shows that $\widehat{\mathcal{B}}_t = O_P(n^{1/6})$ under the alternative hypothesis. Third, the discussion immediately following (23) shows that, under both the null and alternative hypotheses,

$(\widehat{\Upsilon}_t)^{1/2} \xrightarrow{P} (27/c_1^3 \int_0^t \gamma_s^4 ds)^{1/2}$. Therefore,

$$T_n = \left(\frac{\int_0^t \widetilde{\sigma}_s^2 ds}{\sqrt{\frac{27}{c_1^3} \int_0^t \gamma_s^4 ds}} + o_P(1) \right) \times n^{1/5} + O_P(n^{1/6})$$

showing that the divergence rate of the feasible test statistic T_n remains the same as that of the infeasible statistic $\widetilde{T}_n$ under the alternative hypothesis.

Since T_n diverges at a rate $n^{1/5}$ which is faster than that of $\breve{T}_n$ under the alternative hypothesis, we expect that the test T_n should be more powerful than the preliminary test $\breve{T}_n$ when the sample size n is sufficiently large. Let α be the significance level and $z_{1-\alpha}$ the $1-\alpha$ quantile of the standard normal distribution. The asymptotics of the above feasible test are summarized into the following theorem:

Theorem 3. *Suppose that Assumption 1 holds and $1/(6(2-\beta)) < \theta_2 < 1/4$. Under the null hypothesis where $\widetilde{\sigma} \equiv 0$. Then, we have*

$$T_n \xrightarrow{L} N(0, 1)$$

and hence $\lim_{n\to\infty} P(T_n > z_{1-\alpha}|H_0) = \alpha$. Under the alternative hypothesis where $\int_0^t \widetilde{\sigma}_s^2 ds > 0$. Then, we have

$$T_n \asymp_P n^{1/5}, \ T_n \xrightarrow{P} \infty,$$

where $a_n \asymp_P b_n$ means $a_n/b_n = O_P(1)$ and $b_n/a_n = O_P(1)$, and hence $\lim_{n\to\infty} P(T_n > z_{1-\alpha}|H_1) = 1$.

Remark 9. As can be seen from the proof of Theorem 3, the convergence in fact holds for a wider range of θ_2, i.e., $1/(8(2-\beta)) < \theta_2 < 1/4$, under the alternative hypothesis.

2.4. Choices of tuning parameters

We elaborate on the choices of tuning parameters involved in our estimation and testing methods. We first consider the tuning parameter c_3 that appears in the expression for γ_t^2, which is used for estimating the asymptotic variance $\int_0^t \Sigma_s^2 ds$ defined in (13). Recall expression (6) for γ_t^2. Note that σ_ϵ is usually small compared with the spot volatility σ and $\Phi_{22} c_3 \sigma_t^4$ is the dominant term, hence large values of c_3 usually lead to large γ_t^2. Additionally, as can be seen from the expression of the asymptotic variance (13), large values of γ_t^2 lead to large $\int_0^t \Sigma_s^2 ds$. Therefore, a moderate value of c_3 should be used. We adopt a value of $c_3 = 0.1$,[2] which corresponds to $k_n = 242$ ($\approx$ 4 minutes) in the case of the second-by-second data used in our numerical studies. This amounts to using a 4-minute window to mitigate the impact of microstructure noise by pre-averaging. For the choices of c_2 and θ_2, which appear in the threshold $u_n = c_2 n^{-\theta_2}$ (see (5)) for truncating price jumps, we recall the condition $\theta_2 < 1/4$ of Theorems 2 and 3 and, following Aït-Sahalia et al. (2013), set

$$u_n = c_2 \sqrt{\widehat{\sigma^2}} n^{-1/4}$$

with $c_2 = 4$, where $\widehat{\sigma^2}$ is an estimate of the spot variance and we adopt the realized volatility computed using returns that are sampled every 5-minute and spanned one year. As to c_1 that controls the size of blocks over which spot volatilities are estimated, we set $c_1 = 0.4$ in simulation studies and choose it adaptively in empirical studies based on a standard error criterion as follows[3]:

> **Data-driven choice of** c_1. Start estimating the VoV with an initial value $c_{1,0}$ for c_1, and record the associated standard error. Increase c_1 by a step size $c_{1,ss}$, repeat this estimation process and record the resulting standard error. Stop when the absolute value of the *marginal* decrease in the standard error (in percentage terms with respect to the baseline case when $c_1 = c_{1,0}$) drops below a threshold $c_{1,thd}$. Take the last value of c_1 as the final output.

We adopt a fixed $c_1 = 0.4$ in simulation studies because simulations serve the purpose of verifying our theoretical results, where the parameters involved are supposed to be fixed constants. We do not pursue a formal theory of setting tuning parameter values adaptively, as this substantially complicates the theoretical derivations and is of independent interest. We leave this for future research. Nonetheless, our unreported simulation results show that methods based on setting c_1 adaptively offer very similar performance (in terms of the finite-sample bias and root mean square error (RMSE) for estimation, and size and power for the test) to those which use a fixed value of $c_1 = 0.4$.

[2] As discussed in Remark 5, in general, it is difficult to find an optimal choice of c_3. Simulation and empirical results are quite robust to values of $c_3 \in [0.1, 0.2]$.

[3] We implement the data-driven procedure in our empirical study with $(c_{1,0}, c_{1,ss}, c_{1,thd}) = (0.2, 0.1, 5\%)$. The (unreported) numerical results are fairly robust to different choices of these quantities. Based on the feasible CLT Theorem 2, the standard error is computed as $n^{-1/8}(\Xi_t^n)^{1/2}$, where Ξ_t^n is given by (17).

Table 1
Sample bias and RMSE of the volatility of volatility estimates produced by $\widehat{\langle \nu, \nu \rangle}_1$ as defined in (11) under the Heston model with jumps (26) based on 1000 replications.

Sampling intervals	10 s	5 s	2 s	1 s
Bias	−0.0115	−0.0066	−0.0062	−0.0042
RMSE	0.0317	0.0272	0.0241	0.0224

Notes: The target value $\langle \nu, \nu \rangle_1$ in the simulation is approximately 0.1.

3. Simulation study

In this section, we conduct simulation studies to investigate the performance of our estimation and testing methods. The results show that our approaches perform well in finite samples.

3.1. Simulation design

We consider two data generating processes with Poisson jumps for the latent log prices, namely, the Heston stochastic volatility model and a mixed model where the instantaneous volatility process is given by a linear combination of a deterministic component and the square root (volatility) process of Heston model.

The Heston model
The Heston model with Poisson jumps is defined as follows:

$$\begin{cases} dX_t = (\mu^{(1)} - \nu_t/2)dt + \nu_t^{1/2}\, dW_t + dJ_t, \\ d\nu_t = \kappa^{(1)}(\alpha^{(1)} - \nu_t)dt + \xi \nu_t^{1/2}\, d\widetilde{W}_t, \end{cases} \tag{26}$$

where $J_t = \sum_{i=1}^{N_t} z_i$, (z_i) are i.i.d. normal with mean zero and variance σ_J^2, N_t is a Poisson process with intensity λ, and $d\langle W, \widetilde{W}\rangle_t = \rho\, dt$. Taking Aït-Sahalia et al. (2013) as reference, the parameter values are set as $\mu^{(1)} = 0.05$, $\kappa^{(1)} = 10$, $\alpha^{(1)} = 0.1$, $\rho = -0.5$, $\sigma_J = 0.015$, $\lambda = 2$, and $X_0 = \log(40)$. Without further mentioning, we shall always assume $\xi = 1$, whereas in problems related to testing, we consider various values of ξ covering the null and alternative hypotheses. Note that the Feller condition $2\kappa^{(1)}\alpha^{(1)} > \xi^2$ holds for $\xi < 2^{1/2}$ under the above setting of parameter values, ensuring that the spot variance process ν_t is almost surely positive. The initial value ν_0 of ν_t is randomly sampled from its stationary invariant distribution.

The mixed model
The mixed model with Poisson jumps has the following specification:

$$\begin{cases} dX_t = \vartheta(\mu^{(2)} - X_t)dt + \sigma_t dW_t + dJ_t, \\ \sigma_t^2 = (1-\zeta)\sigma_{0,t}^2 + \zeta \nu_t, \\ \sigma_{0,t} = \upsilon(1 + 0.1\cos(2\pi t)), \\ d\nu_t = \kappa^{(1)}(\alpha^{(1)} - \nu_t)dt + \xi \nu_t^{1/2}\, d\widetilde{W}_t, \end{cases} \tag{27}$$

where $J_t = \sum_{i=1}^{N_t} z_i$, (z_i) are i.i.d. normal with mean zero and variance σ_J^2, N_t is a Poisson process with intensity λ, the volatility process σ_t^2 consists of two components, while $\sigma_{0,t}$ has a "U" shape, ν_t is exactly the same as in the Heston model (26). The correlation between the two Brownian motions that drive ν_t and the price process X_t is also set as −0.5. The parameter values for ν_t are the same as those in the Heston model. Other model parameters are set as follows: $\vartheta = 0.5$, $\mu^{(2)} = 4$, $\upsilon = 0.3$, $\sigma_J = 0.015$, $\lambda = 2$, and $X_0 = \log(40)$. We allow $\zeta \in [0, 1]$ to cover the null and alternative hypotheses.

The noise process ϵ_i is taken to be i.i.d. normal with mean zero and variance $\sigma_\epsilon^2 = 0.0001^2$. The terminal time point is chosen to be $T = 1$ year. And the sampling interval is set as one second, meaning that we have $n = 252 \times 6.5 \times 3600$ observations per sample-path. The number of Monte Carlo trials is 1000.

3.2. Estimation results

Table 1 reports the performance of the estimator $\widehat{\langle \nu, \nu \rangle}_1$ under the Heston model with jumps. As expected, the results show that the RMSE decreases as the sampling interval becomes smaller. In particular, the estimator performs well at one-second frequency. In practice, one may use observations at higher frequencies to achieve better accuracy.

Table 2 reports the performance of the estimators $\widehat{\langle \nu, \nu \rangle}_1$ (defined in (11)) and $\widehat{\langle \nu, \nu \rangle}_1^{(4/5)}$ (defined in (21) with $\theta_1 = 4/5$) under the Heston model (26) with $\xi = 0$ and the mixed model (27) with $\zeta = 0$. Note that the true VoV's for the two models are zero. While the first estimator $\widehat{\langle \nu, \nu \rangle}_1$ has a convergence rate of $n^{1/8}$, the second estimator $\widehat{\langle \nu, \nu \rangle}_1^{(4/5)}$ converges at a rate of $n^{1/5}$.

Table 2
Sample mean and RMSE of the estimates produced by $\widehat{\langle \nu, \nu \rangle}_1$ (defined in (11)) and $\widehat{\langle \nu, \nu \rangle}_1^{(4/5)}$ (defined in (21) with $\theta_1 = 4/5$) under the Heston model (26) with $\xi = 0$ and the mixed model (27) with $\zeta = 0$ based on 1000 replications.

		Heston model ($\xi = 0$)				Mixed model (27) ($\zeta = 0$)			
Sampling intervals		10 s	5 s	2 s	1 s	10 s	5 s	2 s	1 s
$\widehat{\langle \nu, \nu \rangle}_1$	mean	0.0000	−0.0001	0.0002	0.0003	0.0010	0.0008	0.0003	0.0004
	RMSE	0.0041	0.0037	0.0034	0.0032	0.0044	0.0040	0.0034	0.0031
$\widehat{\langle \nu, \nu \rangle}_1^{(4/5)}$	mean	−0.0001	−0.0002	0.0000	0.0000	0.0002	0.0002	0.0001	0.0001
	RMSE	0.0015	0.0013	0.0011	0.0010	0.0015	0.0013	0.0010	0.0009

Table 3
Rejection percentages (sizes) of the test using T_n (defined in (25)) based on 1000 replications when data are generated from the Heston model (26) with $\xi = 0$ and the mixed model (27) with $\zeta = 0$.

		Heston model ($\xi = 0$)				Mixed model (27) ($\zeta = 0$)			
Sampling intervals		10 s	5 s	2 s	1 s	10 s	5 s	2 s	1 s
Nominal levels	10%	10.4%	8.2%	10.2%	10.1%	9.6%	9.8%	8.9%	10.3%
	5%	6.2%	5.4%	6.5%	5.8%	5.1%	6.8%	5.6%	5.6%
	2.5%	4.1%	3.5%	3.8%	3.1%	3.3%	4.0%	3.6%	3.7%
	1%	2.2%	2.0%	2.2%	1.8%	2.6%	2.3%	2.0%	1.9%

Notes: The sizes at all significance levels (10%, 5%, 2.5%, 1%) are close to the nominal values.

Table 4
Powers of the test using T_n (defined in (25)) based on 1000 replications when data are generated from the Heston model (26) and the mixed model (27) for different values of ξ and ζ.

		Heston model				Mixed model (27)			
		ξ				ζ			
		0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8
Sampling intervals	10 s	44.0%	97.4%	99.8%	100%	41.2%	97.4%	99.9%	100%
	5 s	47.1%	99.0%	99.9%	100%	49.6%	98.7%	100%	100%
	2 s	62.4%	99.9%	100%	100%	64.7%	99.7%	100%	100%
	1 s	65.1%	99.9%	100%	100%	69.5%	99.8%	100%	100%

Notes: The significance level of the test is 5%. Rejection rates of different scenarios are reported.

We have the following observations from Table 2:

- As the sample size n increases, the RMSE of both $\widehat{\langle \nu, \nu \rangle}$ and $\widehat{\langle \nu, \nu \rangle}^{(4/5)}$ decreases.
- The RMSE of the estimator $\widehat{\langle \nu, \nu \rangle}$ is greater than that of the estimator $\widehat{\langle \nu, \nu \rangle}^{(4/5)}$, confirming the faster rate of convergence of $\widehat{\langle \nu, \nu \rangle}^{(4/5)}$. This suggests that a test statistic based on the latter estimator should be more powerful than that based on the former estimator.

3.3. Test results

We now investigate the performance of the test statistic T_n defined in (25). We first examine the finite-sample size property of the test statistic T_n under the Heston model (26) with $\xi = 0$ and the mixed model (27) with $\zeta = 0$. Table 3 reports the sizes (rejection percentages) for various significance levels. The sizes at all significance levels (10%, 5%, 2.5%, 1%) are close to the nominal values.

We next investigate the power of the above test statistic under both the Heston model and mixed model (27) for various positive values of ξ and ζ. The larger the value of ξ (resp. ζ), the larger the true volatility of volatility $\xi^2 \int_0^1 \nu_s ds$ (resp. $\zeta^2 \int_0^1 \nu_s ds$). Table 4 demonstrates the power of the test (at 5% significance level) based on T_n under both the Heston model and mixed model (27) for various values of ξ and ζ.

We observe from Table 4 the following (under the Heston model (26) and mixed model (27) with the parameter setting specified above):

- With value of $\xi = 0.2$ (or $\zeta = 0.2$), which corresponds to the case when the expectation of the true VoV is $0.2^2 \times 0.1 = 0.004$, it is challenging to separate this case from the null hypothesis. The empirical powers are also reasonably high and increase as the sampling interval shrinks.
- As ξ (or ζ) increases to 0.4, which corresponds to the case when the expectation of the true VoV is $0.4^2 \times 0.1 = 0.016$, the empirical powers are high and close to one when the sampling interval is small.
- Values of $\xi \geq 0.6$ (or $\zeta \geq 0.6$) correspond to cases in which the expectation of the true VoV is 0.036 or larger. The powers are all equal to one when the sampling interval is small.

To summarize, the test T_n is powerful when the data generating process deviates from the null hypothesis.

Table 5
Powers of the test using T_n (defined in (25)) based on 1000 replications when data are generated from the (rough) stochastic volatility model (28) with different values of Hurst parameter H.

H		0.5	0.4	0.3	0.2	0.1
Sampling intervals	10 s	59.2%	89.7%	98.5%	99.9%	100%
	5 s	65.5%	95.3%	99.8%	100%	100%
	2 s	81.3%	98.5%	100%	100%	100%
	1 s	84.0%	98.5%	100%	100%	100%

Notes: The significance level of the test is 5%. Rejection rates of different scenarios are reported.

3.4. Discussion on alternatives beyond diffusions

We document at the beginning of Section 2 that the Itô semimartingale assumption for the volatility process is widely adopted in the literature and we make this assumption throughout the paper. However, given the recent discussions on rough volatility (see, for example, Gatheral et al., 2018), we devote this subsection for such potential alternatives beyond diffusions.

Gatheral et al. (2018) document an interesting finding in their empirical studies that the volatility process may be better described by a process driven by fractional Brownian motion with the rough path. Hence, one might prefer to embed H_1 in a more extensive alternative, for instance, a diffusion process or a process driven by a fractional Brownian motion with Hurst index less than 1/2 (i.e., rough volatility whose quadratic variation does not exist). In this case, the VoV needs to be defined differently according to the Hurst index. However, to the best of our knowledge, an asymptotic theory is not available to estimate the Hurst index under the *stochastic volatility* framework. To formally test the roughness of volatility processes requires an accurate estimation of the Hurst parameter and the corresponding CLT, which is nontrivial. We do not pursue a formal establishment of testing theory about volatility roughness and choose to stay within the popular Itô semimartingale setup to avoid diluting this paper's focus, which is the estimation of VoV and tests based on VoV. Nonetheless, in this subsection, we investigate our test's performance under such a more extensive alternative set, which includes rough volatility, through simulations. Moreover, we also derive the asymptotic behavior of our test statistic under a general rough volatility framework in Appendix B.

A jump–diffusion type stochastic volatility price model where the underlying volatility process is driven by the (rough) fractional Brownian motion can be specified as follows

$$\begin{cases} dX_t = (\mu^{(1)} - \nu_t/2)dt + \nu_t^{1/2}\,dW_t + dJ_t, \\ \nu_t^{1/2} = \exp(\sigma_t^{\nu}), \\ d\sigma_t^{\nu} = \kappa^{(2)}(\alpha^{(2)} - \sigma_t^{\nu})dt + \sigma_H dB_t^H, \end{cases} \tag{28}$$

where $\mu^{(1)}$ and the jump process J_t are set the same as in (26), B_t^H is a fractional Brownian motion with Hurst parameter $H \in (0,1)$ and is independent of W_t. When $H = 1/2$, B_t^H is the classical Brownian motion; when $0 < H < 1/2$, the path of B_t^H is rough and has short memory; when $1/2 < H < 1$, B_t^H has long memory. Gatheral et al. (2018) find that the volatility of the analyzed data can be described by processes driven by the fractional Brownian motion with $0 < H < 1/2$. Taking Gatheral et al. (2018) as reference, the parameter values in this simulation study are set as $\kappa^{(2)} = 5 \times 10^{-4}$, $\alpha^{(2)} = \ln(0.1)/2$, $\sigma_H = 0.4$, $\sigma_0^{\nu} = \alpha^{(2)}$. When $H = 1/2$, the instantaneous variance ν at the long-run mean $\alpha^{(2)}$ of σ_t^{ν} is 0.1 which is comparable with that of the Heston model (26). Again, the noise process ϵ_i is taken to be i.i.d. normal with mean zero and variance $\sigma_\epsilon^2 = 0.0001^2$. The terminal time point is chosen to be $T = 1$ year. And the sampling interval is set as one second, yielding $n = 252 \times 6.5 \times 3600$ observations per sample-path. The number of Monte Carlo trials is 1000.

From the results in Table 5, one can see that the test's power is high under such rough volatility alternatives. However, our test cannot help infer further the roughness of volatility. To distinguish between rough volatility (i.e., driven by a fractional Brownian motion) and non-rough volatility (i.e., driven by a Brownian motion) under the fractional Brownian motion framework, one has to devise a sophisticated testing method that is built on the accurate estimation of Hurst parameter. We leave this for future research.

Remark 10. In addition to this simulation study regarding the power of T_n under rough volatility alternatives, we formally derive its theoretical asymptotic property in Appendix B (see Theorem 4 in Appendix B) under a rough volatility framework that is more general than (28), where we show that T_n diverges at a rate $n^{2(1-H)/5}$ given that $0 < H < 1/2$ and other regularity conditions.

4. Empirical analysis

In this section, we apply the proposed estimator and tests to the high frequency transaction data of SPDR S&P 500 (NYSE Arca: SPY), which is the ETF that tracks the S&P 500 stock market index. With the exception of c_1, we use the same

Table 6
Estimation results based on second-by-second trade price data of SPY during 2004–2013.

Year	VoV	VoV S.E.	VoV/IV	Year	VoV	VoV S.E.	VoV/IV
2004	0.0017	$4e-04$	0.1417	2009	0.0201	0.0089	0.3608
2005	0.0011	$3e-04$	0.1111	2010	0.0179	0.0074	0.716
2006	0.0012	$2e-04$	0.1290	2011	0.0794	0.0506	2.3422
2007	0.011	0.0025	0.0570	2012	0.0011	$3e-04$	0.0894
2008	0.9816	0.3745	8.2975	2013	0.0028	$8e-04$	0.3077

Notes: "VoV" denotes volatility of volatility estimates produced by $\widehat{\langle \nu, \nu \rangle}_1$ (defined in (11)), "VoV S.E." denotes their standard errors computed based on (16) and (17), and "IV" denotes the (yearly) pre-averaging integrated volatility estimates produced by $IV_{0,n}^{PA}$ defined in (3).

Table 7
Realizations of T_n (defined in (25)) and the associated p-values based on second-by-second transaction data of SPY during 2004–2013.

Year	T_n	p-value of T_n	Year	T_n	p-value of T_n
2004	7.5437	0	2009	3.4830	0.0002
2005	1.9517	0.0250	2010	4.3108	0
2006	13.2027	0	2011	9.9531	0
2007	7.4298	0	2012	4.5848	0
2008	12.6177	0	2013	5.7119	0

tuning parameter values as in the simulation studies, i.e., $c_3 = 0.1$ and $u_n = 4(\widehat{\sigma^2})^{1/2} n^{-1/4}$, for estimation and tests. As discussed in Section 2.4, in empirical analyses, we shall choose c_1 adaptively to control the window length for estimating spot volatilities based on a standard error criterion. Moreover, the value of c_1 selected by the data-driven method for the estimation is also adopted for the tests.

4.1. Data

We obtained high frequency data from the NYSE TAQ database. The data set under study spans ten years from Jan 2, 2004, to Dec 31, 2013, covering the 2008 subprime mortgage crisis. We construct the second-by-second price time series by taking the closing price in each one-second interval. The conventional previous-tick method of Zhang (2011) was adopted to deal with those time stamps without price observations. Furthermore, we apply the price validation procedure of Zhou (1996) to remove price bounce-backs. We only consider data during trading hours (from 09:30 to 16:00). Because of the relatively low convergence rate in estimating the VoV, we apply our methods to yearly second-by-second price data. In constructing second-by-second price series that span one year, we remove the overnight returns.

4.2. Estimation results

In this subsection, we apply our estimator to the high frequency data of SPY. Table 6 presents annual estimation results from 2004 to 2013. For each year, we report the estimated VoV and its associated standard error. From Table 6, it is clear that the VoV estimate peaks in 2008, the year in which the subprime mortgage crisis occurred. Moreover, the VoV estimates for SPY in 2004–2006 and 2012–2013 have magnitudes that are very close to zero (ranging from 0.0011 to 0.0028). However, the ratio of VoV estimate to yearly integrated volatility estimate ranges from 8.94% to 30.77% in these years. Hence, it is natural to ask whether the instantaneous volatility processes could be described by Itô processes without diffusion components during these years. Additionally, are the VoV's of the volatility processes during other years (statistically) significantly larger than zero? We shall answer these questions formally in the next subsection.

4.3. Test results

In this subsection, we formally test the null hypothesis that the SPY volatility process can be modeled by a pure drift process.

We report realizations of T_n statistic (defined in (25) with $n^{1/5}$ rate of convergence) and the associated p-values in Table 7. Based on the results, at 5% significance level, the test rejects the null hypothesis for all the ten years under study. Therefore, the conclusion is that pure drift processes are not sufficient for modeling the SPY's volatility processes during these years.

An interesting observation is about year 2011. Notice that based on Table 6, the 95% confidence interval of VoV contains zero. Not surprisingly, the test $\breve{T}_n$ (defined in (20), with $n^{1/8}$ rate of convergence) fails to reject the null hypothesis (unreported). We do obtain a different result with the more powerful test T_n.

4.4. Discussion

We find in empirical analysis that VoV estimates are highly time-varying from year to year and that small (large) VoV estimates are associated with small (large) standard errors. This draws caution in interpreting the point estimates. For instance, "small" VoV estimates do not necessarily imply that pure drift processes are sufficient in modeling volatility processes. The VoV estimates are relatively small for SPY in 2004–2006 and 2012–2013; however, our test (strongly) rejects the zero-VoV null hypothesis.

Additional empirical results (unreported) based on individual stocks such as Johnson & Johnson (NYSE: JNJ), Microsoft Corporation (NASDAQ: MSFT), International Business Machines Corporation (NYSE: IBM), and McDonald's Corporation (NYSE: MCD) demonstrate similar features.

5. Concluding remarks

This paper is the first to derive inference results for the VoV based on noisy high frequency data with jumps. We first use a threshold pre-averaging method to estimate the spot volatilities and then construct an estimator of the VoV through proper bias correction to the realized VoV. The corresponding CLT and feasible CLT are established with a convergence rate of $n^{1/8}$. We illustrate the applications of our feasible CLT in estimating and testing VoV. We also construct a more powerful test for the null hypothesis that VoV is zero and show that the test statistic has a divergence rate of $n^{1/5}$ under the alternative hypothesis that a diffusion component is present in the volatility process.

We show in simulations that the proposed VoV estimator and the test statistic perform well in finite samples. We also apply the VoV estimator and tests to the second-by-second trade price data of SPY during 2004–2013 and individual stocks. The empirical results show strong evidence of non-zero VoV.

One potential direction for improving VoV estimation accuracy is implementing the idea of Estimated-price Realized Volatility (Li et al., 2016). This achieves higher efficiency with trading information taken into account and can be applied to data of higher frequencies (e.g., tick-by-tick). We leave this for future research.

Acknowledgments

Li acknowledges support from the Research Grants Council, Hong Kong SAR GRF16518716, GRF16503419, GRF16502118 and T31-604/18N as well as the National Natural Science Foundation of China NSFC19BM03. Liu acknowledges support from the National Natural Science Foundation of China (11501503, 71971118); National Social Science Fund Project of China (19BTJ035); Qinglan Project of Jiangsu Province, China; Natural Science Foundation of Jiangsu Province of China (BK20181417); and Jiangsu Province College Science Key Foundation, China (17KJA110001). Zhang acknowledges support from the National Natural Science Foundation of China (71871132 and 91546202).

Appendix A. The proofs for the main results

In this appendix, we provide proofs of the main results and some supplementary theoretical results. We use K to denote a constant that may change from line to line, and use K_p when we emphasize that it depends on an additional parameter p. To simplify notation, we denote

$$\delta_n = 1/n, \quad \Delta_n = l_n\delta_n = l_n/n.$$

E_i^n denotes the conditional expectation with respect to the σ-field $\mathcal{F}_{i\delta_n}$. Some of the following notation follows from Jacod et al. (2009). For each n, we define the function

$$g_n(s) = \sum_{j=1}^{k_n-1} g_j^n 1_{((j-1)\delta_n, j\delta_n]}(s),$$

which is equal to zero for $s > (k_n-1)\delta_n$ and $s \le 0$, and is bounded uniformly in n. We introduce the processes

$$X(n,s)_t = \int_0^{s+t} b_u g_n(u-s)du + \int_0^{s+t} \sigma_u g_n(u-s)dW_u, \tag{A.1}$$

$$C(n,s)_t = \int_0^{s+t} \sigma_u^2 g_n(u-s)^2 du, \quad C_t = \int_0^t \sigma_s^2 ds, \quad c_i^n = \sum_{j=1}^{k_n-1} (g_j^n)^2 \Delta_{i+j}^n C,$$

and note that the processes $X(n,s)_t$ and $C(n,s)_t$ vanish for $t \le 0$ and are constant for $t \ge (k_n-1)\delta_n$. Moreover, we define the following two processes

$$B(n,s)_t = \int_0^{s+t} X(n,s)_u b_u g_n(u-s)du, \quad M(n,s)_t = \int_0^{s+t} X(n,s)_u \sigma_u g_n(u-s)dW_u,$$

One readily sees that Itô's formula implies

$$X(n,s)^2 = 2B(n,s) + C(n,s) + 2M(n,s). \tag{A.2}$$

Setting $X_t^c = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s$, we can rewrite $X_t = X_t^c + J_t$ and $\overline{X^c}_i^n = X(n, i\delta_n)_{k_n\delta_n}$. By a standard localization procedure, the local boundedness of the processes $b, \widetilde{b}, \overline{b}, \sigma, \widetilde{\sigma}$ and $\overline{\sigma}$ can be replaced by a boundedness (uniformly in (ω, t)) condition; see, for example, Aït-Sahalia and Jacod (2014). Hence, in the following, we assume that $b, \widetilde{b}, \overline{b}, \sigma, \widetilde{\sigma}$ and $\overline{\sigma}$ are bounded. Similarly, following Jacod and Protter (2012), we also assume that X is bounded.

Proof of Proposition 1. We only prove (7), and the rest results can be derived using similar arguments. Note that $Y = X^c + \epsilon + J$. By (3), we obtain

$$\begin{aligned} l_n^{\frac{1}{2}} n^{-\frac{1}{4}} (\widehat{\nu_t} - \nu_t) = {} & l_n^{\frac{1}{2}} n^{-\frac{1}{4}} \left(\frac{n}{l_n} \left(\frac{1}{\psi_2 k_n} \sum_{j=1}^{l_n-k_n+1} (\overline{X^c+\epsilon}_{i+j}^n)^2 - \frac{\psi_1}{2\psi_2 k_n^2} \sum_{j=0}^{l_n} (\Delta_{i+j}^n (X^c+\epsilon))^2 \right) - \nu_t \right) \\ & + l_n^{\frac{1}{2}} n^{-\frac{1}{4}} \left(\frac{n}{\psi_2 k_n l_n} \sum_{j=1}^{l_n-k_n+1} [(\overline{Y}_{i+j}^n)^2 1_{|\overline{Y}_{i+j}^n|<u_n} - (\overline{X^c+\epsilon}_{i+j}^n)^2] \right. \\ & \left. - \frac{\psi_1 n}{2\psi_2 l_n k_n^2} \sum_{j=0}^{l_n} [(\Delta_{i+j}^n Y)^2 - (\Delta_{i+j}^n (X^c+\epsilon))^2] \right). \end{aligned}$$

Following the result of Aït-Sahalia and Jacod (2014) on page 296 and (3), we obtain that the first term $\xrightarrow{L-s} Z_t + c_1 Z_t'$. Hence, it remains to show that the second term is of $o_P(1)$. It thus suffices to show that

$$\frac{n^{3/4}}{k_n l_n^{1/2}} \sum_{j=1}^{l_n-k_n+1} (\overline{J}_{i+j}^n)^2 1_{|\overline{Y}_{i+j}^n|<u_n} \xrightarrow{P} 0, \tag{A.3}$$

$$\frac{n^{3/4}}{k_n l_n^{1/2}} \sum_{j=1}^{l_n-k_n+1} (\overline{X^c+\epsilon}_{i+j}^n)^2 1_{|\overline{Y}_{i+j}^n|\geq u_n} \xrightarrow{P} 0, \tag{A.4}$$

$$\frac{n^{3/4}}{k_n^2 l_n^{1/2}} \sum_{j=1}^{l_n} (\Delta_{i+j}^n J)^2 \xrightarrow{P} 0, \tag{A.5}$$

$$\frac{n^{3/4}}{k_n^2 l_n^{1/2}} \sum_{j=1}^{l_n} (\Delta_{i+j}^n (X^c+\epsilon))^2 \xrightarrow{P} 0. \tag{A.6}$$

We only treat (A.3) and (A.4), while (A.5) and (A.6) follows easily.

For the result (A.4), noticing that $E_i^n|\overline{X^c+\epsilon}_i^n|^r \leq Kn^{-r/4}$ for $r > 0$, $E_i^n|\overline{J}_i^n| \leq Kn^{-1/2}$ and $\theta_2 < 1/4$, by Hölder's inequality, we have

$$\begin{aligned} & E_i^n(\overline{X^c+\epsilon}_i^n)^2 1_{|\overline{Y}_i^n|\geq u_n} \leq (E_i^n(\overline{X^c+\epsilon}_i^n)^{2a})^{\frac{1}{a}} (E_i^n 1_{|\overline{Y}_i^n|\geq u_n})^{\frac{1}{b}} \\ & \leq Kn^{-1/2} (E_i^n 1_{|\overline{X^c+\epsilon}_i^n|\geq u_n/2} + E_i^n 1_{|\overline{J}_i^n|\geq u_n/2})^{\frac{1}{b}} \leq Kn^{-1/2} \left(E_i^n \left(\frac{|\overline{X^c+\epsilon}_i^n|}{u_n} \right)^r + E_i^n \frac{|\overline{J}_i^n|}{u_n} \right)^{\frac{1}{b}} \\ & \leq Kn^{-1/2} \left(n^{(\theta_2-1/4)r} + n^{\theta_2-1/2} \right)^{\frac{1}{b}}, \end{aligned}$$

for $1/a + 1/b = 1$ with $a, b > 0$ and $r > 0$. Let b be sufficiently close to 1. Then the arbitrariness of r and $\theta_2 < 1/4$ imply that result (A.4) holds.

For the result (A.3), similar to result (A.4) of Jing et al. (2014), we have

$$E_i^n|\overline{J}_i^n \wedge u_n| \leq Kn^{-1/2} u_n^{1-s}, \quad \text{for } \beta < s < 1. \tag{A.7}$$

Note that

$$1_{|\overline{Y}_i^n|\leq u_n} \leq 1_{|\overline{J}_i^n|\leq 2u_n} + 1_{|\overline{X^c+\epsilon}_i^n|>u_n}, \tag{A.8}$$

similar to the result (A.4), so it is sufficient to show that

$$\frac{n^{3/4}}{k_n l_n^{1/2}} \sum_{j=1}^{l_n} (\overline{J}_{i+j}^n)^2 1_{|\overline{J}_{i+j}^n|<2u_n} \xrightarrow{P} 0. \tag{A.9}$$

By (A.7), we have

$$E_i^n|\frac{n^{3/4}}{k_n l_n^{1/2}}\sum_{j=1}^{l_n}(\bar{J}_{i+j}^n)^2 1_{|\bar{J}_{i+j}^n|<2u_n}| \le Kn^{\frac{5}{8}}\max_{i\le j\le i+l_n}E_i^n(\bar{J}_j^n)^2 1_{|\bar{J}_j^n|<2u_n} \le Kn^{\frac{5}{8}}k_n\delta_n u_n^{2-s} \le Kn^{\frac{1}{8}-\theta_2(2-s)},$$

and so the condition $\theta_2 > 1/(8(2-\beta))$ implies (A.9). □

We need an approximation lemma to prove Proposition 2. The lemma and Proposition 2 involve the following notation that are used to rewrite $\widehat{\nu}_{(i+m)\Delta_n} - \widehat{\nu}_{i\Delta_n}$:

$$\eta(1,m)_i^n = \frac{1}{\psi_2 k_n l_n\delta_n}\sum_{j=0}^{l_n-k_n-1}[c_{(i+m)l_n+j}^n - c_{il_n+j}^n],$$

$$\eta(2,m)_i^n = \frac{1}{\psi_2 k_n l_n\delta_n}\sum_{j=0}^{l_n-k_n-1}[(\overline{X^c}_{(i+m)l_n+j}^n)^2 - c_{(i+m)l_n+j}^n - (\overline{X^c}_{il_n+j}^n)^2 + c_{il_n+j}^n],$$

$$\eta(3,m)_i^n = \frac{2}{\psi_2 k_n l_n\delta_n}\sum_{j=0}^{l_n-k_n-1}[\overline{X^c}_{(i+m)l_n+j}^n\bar{\epsilon}_{(i+m)l_n+j}^n - \overline{X^c}_{il_n+j}^n\bar{\epsilon}_{il_n+j}^n],$$

$$\eta(4,m)_i^n = \frac{1}{\psi_2 k_n l_n\delta_n}\sum_{j=0}^{l_n-k_n-1}[(\bar{\epsilon}_{(i+m)l_n+j}^n)^2 - (\bar{\epsilon}_{il_n+j}^n)^2],$$

$$\eta(5,m)_i^n = -\frac{1}{2\psi_2 k_n^2 l_n\delta_n}\sum_{j=0}^{l_n}[(\Delta_{(i+m)l_n+j}^n Y)^2 - (\Delta_{il_n+j}^n Y)^2],$$

$$\eta(6,m)_i^n = \frac{-1}{\psi_2 k_n l_n\delta_n}\sum_{j=0}^{l_n-k_n-1}[(\overline{X^c+\epsilon}_{(i+m)l_n+j}^n)^2 1_{|\overline{Y}_{(i+m)l_n+j}^n|\ge u_n} - (\overline{X^c+\epsilon}_{il_n+j}^n)^2 1_{|\overline{Y}_{il_n+j}^n|\ge u_n}],$$

$$\eta(7,m)_i^n = \frac{1}{\psi_2 k_n l_n\delta_n}\sum_{j=0}^{l_n-k_n-1}[(\bar{J}_{(i+m)l_n+j}^n)^2 1_{|\overline{Y}_{(i+m)l_n+j}^n|<u_n} - (\bar{J}_{il_n+j}^n)^2 1_{|\overline{Y}_{il_n+j}^n|<u_n}],$$

$$\eta(8,m)_i^n = \frac{2}{\psi_2 k_n l_n\delta_n}\sum_{j=0}^{l_n-k_n-1}[\overline{X^c+\epsilon}_{(i+m)l_n+j}^n\bar{J}_{(i+m)l_n+j}^n 1_{|\overline{Y}_{(i+m)l_n+j}^n|<u_n} - \overline{X^c+\epsilon}_{il_n+j}^n\bar{J}_{il_n+j}^n 1_{|\overline{Y}_{il_n+j}^n|<u_n}].$$

Then, we can rewrite the increment $\widehat{\nu}_{(i+m)\Delta_n} - \widehat{\nu}_{i\Delta_n}$ as follows:

$$\widehat{\nu}_{(i+m)\Delta_n} - \widehat{\nu}_{i\Delta_n} = \sum_{j=1}^{8}\eta(j,m)_i^n. \tag{A.10}$$

We define the approximate counterparts of $\{\eta(j,m)_i^n\}_{1\le j\le 3}$ as follows:

$$\breve{\eta}(1,m)_i^n = \frac{\widetilde{\sigma}_{i\Delta_n}}{\Delta_n}\int_0^{\Delta_n}(\widetilde{W}_{(i+m)\Delta_n+s} - \widetilde{W}_{i\Delta_n+s})ds,$$

$$\breve{\eta}(2,m)_i^n = \frac{2\sigma_{i\Delta_n}^2}{\psi_2 k_n\Delta_n}\sum_{j=0}^{l_n-k_n-1}(\breve{\eta}(2)_{(i+m)l_n+j}^n - \breve{\eta}(2)_{il_n+j}^n),$$

$$\breve{\eta}(3,m)_i^n = \frac{2\sigma_{i\Delta_n}}{\psi_2 k_n\Delta_n}\sum_{j=0}^{l_n-k_n-1}(\overline{W}_{(i+m)l_n+j}^n\bar{\epsilon}_{(i+m)l_n+j}^n - \overline{W}_{il_n+j}^n\bar{\epsilon}_{il_n+j}^n),$$

where

$$\breve{\eta}(2)_i^n = \int_{i\delta_n}^{(i+k_n)\delta_n}\int_{i\delta_n}^{u}g_n(s-i\delta_n)dW_s g_n(u-i\delta_n)dW_u.$$

The following lemma provides estimates for the bounds of relevant approximation errors and quantities.

Lemma 1. *Suppose that Assumption 1 holds and $r\le 4$. Then,*

$$\left.\begin{array}{l} E_{il_n}^n|\eta(2,m)_i^n - \breve{\eta}(2,m)_i^n|^r \le K_r(k_n\delta_n)^{\frac{r}{2}}, \\ E_{il_n}^n|\eta(3,m)_i^n - \breve{\eta}(3,m)_i^n|^r \le K_r(k_n^{-1})^{\frac{r}{2}}, \end{array}\right\} \tag{A.11}$$

and

$$\left.\begin{array}{ll} E^n_{il_n}|\eta(1,m)^n_i|^r \le K_r(l_n\delta_n)^{\frac{r}{2}}, & E^n_{il_n}|\eta(2,m)^n_i|^r \le K_r(k_n/l_n)^{\frac{r}{2}}, \\ E^n_{il_n}|\eta(3,m)^n_i|^r \le K_r(k_nl_n\delta_n)^{-\frac{r}{2}}, & E^n_{il_n}|\eta(4,m)^n_i|^r \le K_r(k_n^3l_n\delta_n^2)^{-\frac{r}{2}}, \\ E^n_{il_n}|\eta(5,m)^n_i|^r \le K_r(k_n^4l_n\delta_n^2)^{-\frac{r}{2}}, & \end{array}\right\} \tag{A.12}$$

where inequalities (A.12) also hold when $\eta(j,m)^n_i$ is replaced by $\breve{\eta}(j,m)^n_i$ for $j=1,2,3$. *Suppose that Assumption 2 holds and $r \le 4$, then*

$$E^n_{il_n}|\eta(1,m)^n_i - \breve{\eta}(1,m)^n_i|^r \le K_r(l_n\delta_n + k_n/l_n)^r. \tag{A.13}$$

Proof. We only prove the case of $r=2$. For the result of (A.13), we have

$$\begin{aligned} \eta(1,m)^n_i &= \frac{\sum_{j=1}^{k_n-1}(g^n_j)^2}{\psi_2 k_n\Delta_n}\int_0^{\Delta_n}(\sigma^2_{(i+m)\Delta_n+s} - \sigma^2_{i\Delta_n+s})ds + O_P\left(\frac{1}{k_n\Delta_n}\right) \\ &= \frac{1}{\Delta_n}\int_0^{\Delta_n}(\sigma^2_{(i+m)\Delta_n+s} - \sigma^2_{i\Delta_n+s})ds + O_P\left(\frac{1}{k_n\Delta_n}\right), \end{aligned} \tag{A.14}$$

where we use the fact that $\sum_{j=1}^{k_n-1}(g^n_j)^2 = \psi_2 k_n + O(1)$. Further,

$$\begin{aligned} \eta(1,m)^n_i - \breve{\eta}(1,m)^n_i &= \frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+m)\Delta_n+s}(\widetilde{\sigma}_u - \widetilde{\sigma}_{i\Delta_n})d\widetilde{W}_u ds \\ &\quad + \frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+m)\Delta_n+s}\widetilde{b}_u du ds + O_P\left(\frac{1}{k_n\Delta_n}\right). \end{aligned}$$

Then, it is sufficient to show that the result (A.13) holds for the above two parts. The boundedness of $\widetilde{b}$ suggests that

$$E^n_{il_n}\left|\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+m)\Delta_n+s}\widetilde{b}_u du ds\right| \le K\Delta_n.$$

By Hölder's inequality, Burkholder–Davis–Gundy inequality and Assumption 2, we have

$$E^n_{il_n}\left(\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+m)\Delta_n+s}(\widetilde{\sigma}_u - \widetilde{\sigma}_{i\Delta_n})d\widetilde{W}_u ds\right)^2 \le K\Delta_n^2.$$

Hence, we obtain the result (A.13).

For the first result of (A.11), note that from (A.2), it is sufficient to show that

$$E^n_{il_n}\left|\frac{1}{k_n\Delta_n}\sum_{j=1}^{l_n-k_n+1}[M(n,((i+m)l_n+j)\delta_n)_{k_n\delta_n} - \sigma^2_{i\Delta_n}\breve{\eta}(2)^n_{(i+m)l_n+j}]\right| \le K(k_n\delta_n)^{\frac{1}{2}}, \tag{A.15}$$

$$E^n_{il_n}\left|\frac{1}{k_n\Delta_n}\sum_{j=1}^{l_n-k_n+1}[M(n,(il_n+j)\delta_n)_{k_n\delta_n} - \sigma^2_{i\Delta_n}\breve{\eta}(2)^n_{il_n+j}]\right| \le K(k_n\delta_n)^{\frac{1}{2}}, \tag{A.16}$$

$$E^n_{il_n}\left|\frac{1}{k_n\Delta_n}\sum_{j=1}^{l_n-k_n+1}[B(n,((i+m)l_n+j)\delta_n)_{k_n\delta_n} - B(n,(il_n+j)\delta_n)_{k_n\delta_n}]\right| \le K(k_n\delta_n)^{\frac{1}{2}}. \tag{A.17}$$

The boundedness of b,σ,X implies (A.17). For the first two results (A.15) and (A.16), it is sufficient to prove the second inequality (A.16), as the first inequality can be similarly obtained. By (A.1) and the property of Itô integral, we have

$$\begin{aligned} &E^n_i\left(M(n,i\delta_n)_{k_n\delta_n} - \sigma^2_{i\delta_n}\breve{\eta}(2)^n_i\right)^2 = E^n_i\left(\int_{i\delta_n}^{(i+k_n)\delta_n}\int_{i\delta_n}^{u} b_s g_n(s-i\delta_n)\sigma_u g_n(u-i\delta_n)ds dW_u\right. \\ &\left.+\int_{i\delta_n}^{(i+k_n)\delta_n}\int_{i\delta_n}^{u}(\sigma_s\sigma_u - \sigma^2_{i\delta_n})g_n(s-i\delta_n)g_n(u-i\delta_n)dW_s dW_u\right)^2 \\ &\le K\left(E^n_i\int_{i\delta_n}^{(i+k_n)\delta_n}\left(\int_{i\delta_n}^{u}|b_s\sigma_u|ds\right)^2 du + E^n_i\int_{i\delta_n}^{(i+k_n)\delta_n}\int_{i\delta_n}^{u}(\sigma_s\sigma_u - \sigma^2_{i\delta_n})^2 ds du\right) \\ &\le K(k_n\delta_n)^3, \end{aligned} \tag{A.18}$$

where the fact that σ is an Itô diffusion process is used in the last inequality. Note that $E^n_{il_n}(\sigma^2_{(il_n+j_1)\delta_n} - \sigma^2_{il_n\delta_n})\breve{\eta}(2)^n_{il_n+j_1}(\sigma^2_{(il_n+j_2)\delta_n} - \sigma^2_{il_n\delta_n})\breve{\eta}(2)^n_{il_n+j_2} = 0$ for $|j_1 - j_2| > k_n$, we have

$$E^n_{il_n}\left(\frac{1}{k_n\Delta_n}\sum_{j=1}^{l_n-k_n+1}(\sigma^2_{(il_n+j)\delta_n} - \sigma^2_{il_n\delta_n})\breve{\eta}(2)^n_{il_n+j}\right)^2$$
$$\leq \frac{Kl_nk_n}{(k_nl_n\delta_n)^2}\max_{1\leq j\leq l_n} E^n_{il_n}\left((\sigma^2_{(il_n+j)\delta_n} - \sigma^2_{il_n\delta_n})\breve{\eta}(2)^n_{il_n+j}\right)^2 \leq \frac{Kl_nk_n}{(k_nl_n\delta_n)^2}l_n\delta_n(k_n\delta_n)^2 = Kk_n\delta_n,$$

hence, by (A.18) and Cauchy–Schwarz inequality, we prove the result (A.16).

For the second result of (A.11), it is enough to show

$$E^n_{il_n}\left(\frac{1}{k_n\Delta_n}\sum_{j=1}^{l_n-k_n+1}(\overline{X^c}^n_{il_n+j} - \sigma_{i\Delta_n}\overline{W}^n_{il_n+j})\overline{\epsilon}^n_{il_n+j}\right)^2 \leq Kk_n^{-1},$$

by inequality (5.34) of Jacod et al. (2009), it is not hard to show the above inequality.

It is not difficult to show that the results in (A.12) hold. Hence, we have proved Lemma 1. □

We are now ready to give the proof for Proposition 2.

Proof of Proposition 2. Recall that $\theta_1 = 3/4$ and the notations before Lemma 1. By straightforward calculations, we obtain

$$\left.\begin{aligned}
&E^n_{il_n}(\breve{\eta}(1,m)^n_i)^2 = (m - \frac{1}{3})\widetilde{\sigma}^2_{i\Delta_n}\Delta_n,\\
&E^n_{il_n}(\breve{\eta}(2,m)^n_i)^2 = \frac{8\Phi_{22}}{\psi_2^2}\sigma^4_{i\Delta_n}\frac{k_n}{l_n} + O_P\left(\frac{k_n^2}{l_n^2}\right),\\
&E^n_{il_n}(\breve{\eta}(3,m)^n_i)^2 = \frac{16\Phi_{12}}{\psi_2^2}\sigma^2_{i\Delta_n}\sigma^2_\epsilon\frac{1}{k_n\Delta_n} + O_P\left(\frac{1}{l_n\Delta_n}\right),\\
&E^n_{il_n}(\eta(4,m)^n_i)^2 = \frac{8\Phi_{11}}{\psi_2^2}\sigma^4_\epsilon\frac{1}{k_n^3l_n\delta_n^2} + O_P\left(\frac{1}{k_n^2l_n^2\delta_n}\right),\\
&E^n_{il_n}(\eta(5,m)^n_i)^2 = \frac{2}{\psi_2^2}\sigma^4_\epsilon\frac{1}{k_n^4l_n\delta_n^2} + O_P\left(\frac{1}{k_n^4l_n\delta_n}\right).
\end{aligned}\right\} \tag{A.19}$$

For convenience, we write $\breve{\eta}(j,m)^n_i = \eta(j,m)^n_i$ for $j = 4,\ldots,8$, and then

$$\begin{aligned}
V(\nu;\theta_1,m)^n_t &= \sum_{i=1}^{[t/\Delta_n]-m-1}(\widehat{\nu}_{(i+m)\Delta_n} - \widehat{\nu}_{i\Delta_n})^2 = \sum_{i=1}^{[t/\Delta_n]-m-1}\left(\sum_{j=1}^{8}\eta(j,m)^n_i\right)^2\\
&= \sum_{i=1}^{[t/\Delta_n]-m-1}\sum_{j_1=1}^{8}(\breve{\eta}(j_1,m)^n_i)^2 + \sum_{i=1}^{[t/\Delta_n]-m-1}\sum_{j_1\neq j_2}\breve{\eta}(j_1,m)^n_i\breve{\eta}(j_2,m)^n_i\\
&\quad + \sum_{i=1}^{[t/\Delta_n]-m-1}\sum_{j_1,j_2=1}^{8}(\eta(j_1,m)^n_i - \breve{\eta}(j_1,m)^n_i)(\eta(j_2,m)^n_i - \breve{\eta}(j_2,m)^n_i)\\
&\quad + 2\sum_{i=1}^{[t/\Delta_n]-m-1}\sum_{j_1,j_2=1}^{8}(\eta(j_1,m)^n_i - \breve{\eta}(j_1,m)^n_i)\breve{\eta}(j_2,m)^n_i.
\end{aligned}$$

The results in (A.19) imply that

$$\sum_{i=1}^{[t/\Delta_n]-m-1}\sum_{j=1}^{4}E^n_{il_n}(\breve{\eta}(j,m)^n_i)^2 \xrightarrow{P} \left(m - \frac{1}{3}\right)\int_0^t\widetilde{\sigma}^2_s ds + \frac{2}{c_1^2}\int_0^t\gamma^2_s ds.$$

By the Cauchy–Schwarz inequality, then it is enough to show that

$$\sum_{i=1}^{[t/\Delta_n]-m-1}\sum_{j=1}^{4}\left((\breve{\eta}(j,m)^n_i)^2 - E^n_i(\breve{\eta}(j,m)^n_i)^2\right) \xrightarrow{P} 0, \tag{A.20}$$

$$\sum_{i=1}^{[t/\Delta_n]-m-1} (\breve{\eta}(j,m)_i^n)^2 \xrightarrow{P} 0, \quad j=5,\ldots,8, \tag{A.21}$$

$$\sum_{i=1}^{[t/\Delta_n]-m-1} \breve{\eta}(j_1,m)_i^n \breve{\eta}(j_2,m)_i^n \xrightarrow{P} 0, \quad j_1,j_2=1,\ldots,8, j_1\neq j_2, \tag{A.22}$$

$$\sum_{i=1}^{[t/\Delta_n]-m-1} \sum_{j_1=1}^{3} (\eta(j_1,m)_i^n - \breve{\eta}(j_1,m)_i^n)^2 \xrightarrow{P} 0. \tag{A.23}$$

For (A.20), by Lemma 2.2.11 of Jacod and Protter (2012), it is sufficient to prove

$$\sum_{i=1}^{[t/\Delta_n]-m-1} E_{il_n}^n (\breve{\eta}(j,m)_i^n)^4 \xrightarrow{P} 0, \quad j=1,2,\ldots,4. \tag{A.24}$$

For (A.22), the Cauchy–Schwarz inequality implies that it is sufficient to show (A.21) and to show that (A.22) holds for $j_1,j_2=1,\ldots,4, j_1\neq j_2$. For result (A.22) with $j_1,j_2=1,\ldots,4, j_1\neq j_2$, we have

$$E\left(\sum_{i=1}^{[t/\Delta_n]-m-1} \breve{\eta}(j_1,m)_i^n \breve{\eta}(j_2,m)_i^n\right)^2 = \sum_{i=1}^{[t/\Delta_n]-m-1} E\left(\breve{\eta}(j_1,m)_i^n \breve{\eta}(j_2,m)_i^n\right)^2.$$

Then, by Hölder's inequality, it is sufficient to show (A.24). For (A.23), Lemma 2.2.10 of Jacod and Protter (2012) implies that it is sufficient to prove

$$\sum_{i=1}^{[t/\Delta_n]-m-1} E_{il_n}^n (\eta(j,m)_i^n - \breve{\eta}(j,m)_i^n)^2 \xrightarrow{P} 0, \quad j=1,2,3. \tag{A.25}$$

Lemma 1 suggests that arguments (A.24) and (A.25) hold for $j=2,3$. For $j=1$ in (A.25), under Assumption 1, by the proof of (A.13), we have

$$E_{il_n}^n (\eta(j,m)_i^n - \breve{\eta}(j,m)_i^n)^2 \leq K\Delta_n^2 + K\Delta_n E_{il_n}^n (\sup_{s,t\in[i\Delta_n,(i+m)\Delta_n]} |\widetilde{\sigma}_s - \widetilde{\sigma}_t|)^2.$$

Then by Lemma 5.4 of Jacod et al. (2009), we have (A.25) for $j=1$. Note that Lemma 1 implies that (A.21) holds when $j=5$; hence, it remains to show (A.21) for $j=6,7,8$.

For $j=6$ in (A.21), similar to the result (A.4), it is not hard to obtain that (A.21) holds for $j=6$. For $j=7$ in (A.21), by (A.8), similar to the case $j=6$ in (A.21), so it is sufficient to show that

$$\sum_{i=1}^{[t/\Delta_n]-m-1} (\breve{\eta}(7,m)_i^n)^2 \xrightarrow{P} 0, \tag{A.26}$$

where

$$\breve{\eta}(7,m)_i^n = \frac{1}{\psi_2 k_n l_n \delta_n} \sum_{j=0}^{l_n-k_n-1} [(\bar{J}^n_{(i+m)l_n+j})^2 1_{|\bar{J}^n_{(i+m)l_n+j}|<2u_n} + (\bar{J}^n_{il_n+j})^2 1_{|\bar{J}^n_{il_n+j}|<2u_n}].$$

By (A.7), we have

$$E_{il_n}^n |\breve{\eta}(7,m)_i^n| \leq \frac{K}{k_n\delta_n} \max_{il_n\leq j\leq (i+m+1)l_n} E_{il_n}^n (\bar{J}^n_j)^2 1_{|\bar{J}^n_j|<2u_n} \leq \frac{K}{k_n\delta_n} k_n\delta_n u_n^{2-s} \leq K n^{-\theta_2(2-s)}, \tag{A.27}$$

and so the condition $\theta_2 > 1/(8(2-\beta))$ implies (A.26). For $j=8$ in (A.21), by the result for $j=7$, Hölder's inequality suggests that the result holds provided that $\theta_2 > 1/(8(2-\beta))$. □

Remark 11. By very similar arguments to that of the proof for Proposition 2, we can show that the following general results hold under the same Assumption 1. That is, if $1/2<\theta_1<1$ and $(1-\theta_1)/(4-2\beta)<\theta_2<1/4$, then

$$V(\nu;\theta_1,m)_t^n - \frac{2n^{\frac{3}{2}-2\theta_1}}{c_1^2}\int_0^t \gamma_s^2 ds \xrightarrow{P} (m-1/3)\int_0^t \widetilde{\sigma}_s^2 ds. \tag{A.28}$$

Furthermore, suppose that $\widetilde{\sigma}\equiv 0$, $\theta_1=5/6$, and $(1-\theta_1)/(2-\beta)<\theta_2<1/4$, then

$$n^{1-\theta_1} V(\nu;\theta_1,m)_t^n \xrightarrow{P} m^2 c_1 \int_0^t \widetilde{b}_s^2 ds + \frac{2}{c_1^2}\int_0^t \gamma_s^2 ds. \tag{A.29}$$

The convergence results in (A.28) and (A.29) are useful in the construction of our test for whether drift processes are sufficient in describing volatility processes.

To prove Theorem 1, we follow the same arguments as in the proof of Theorem 2.6 of Vetter (2015). Some additional notation and a lemma are needed. Let p be an arbitrary positive integer. Define

$$J_n(p) = [([t/\Delta_n] - 3)/(2(p+2))], \quad a_q(p) := 2(q-1)(p+2), \quad b_q(p) := a_q(p) + 2p, \tag{A.30}$$

for $q = 1, 2, \ldots, J_n(p)$, where $J_n(p)$ is the number of big blocks with common time span $2(p+2)\Delta_n$. Denote $s_q := a_q(p)l_n$. Recall the definitions of $\eta(j,m)_i^n$ right before Lemma 1 for $j = 1, \ldots, 4$. We define their approximate counterparts as follows (with $m = 1$): for $a_q(p) \le i < a_{q+1}(p)$, set

$$\begin{aligned}
\widetilde{\eta}(1,m)_i^n &= \frac{\widetilde{\sigma}_{a_q(p)\Delta_n}}{\Delta_n}\int_0^{\Delta_n}(\widetilde{W}_{(i+m)\Delta_n+s} - \widetilde{W}_{i\Delta_n+s})ds,\\
\widetilde{\eta}(2,m)_i^n &= \frac{2\sigma^2_{a_q(p)\Delta_n}}{\psi_2 k_n\Delta_n}\sum_{j=0}^{l_n-k_n-1}(\breve{\eta}(2)^n_{(i+m)l_n+j} - \breve{\eta}(2)^n_{il_n+j}),\\
\widetilde{\eta}(3,m)_i^n &= \frac{2\sigma_{a_q(p)\Delta_n}}{\psi_2 k_n\Delta_n}\sum_{j=0}^{l_n-k_n-1}(\overline{W}^n_{(i+m)l_n+j}\overline{\epsilon}^n_{(i+m)l_n+j} - \overline{W}^n_{il_n+j}\overline{\epsilon}^n_{il_n+j}).
\end{aligned}$$

Based on the above notation, we further define

$$\begin{aligned}
&\widetilde{U}(1,1)_i^n = (\widetilde{\eta}(1,1)_i^n)^2 - \frac{2}{3}\widetilde{\sigma}^2_{a_q(p)\Delta_n}\Delta_n, \quad \widetilde{U}(2,2)_i^n = (\widetilde{\eta}(2,1)_i^n)^2 - \frac{8c_2\Phi_{22}}{(\psi_2c_1)^2}\sigma^4_{a_q(p)\Delta_n}\Delta_n,\\
&\widetilde{U}(3,3)_i^n = (\widetilde{\eta}(3,1)_i^n)^2 - \frac{16\Phi_{12}\sigma_\epsilon^2}{c_2(\psi_2c_1)^2}\sigma^2_{a_q(p)\Delta_n}\Delta_n, \quad \widetilde{U}(4,4)_i^n = (\widetilde{\eta}(4,1)_i^n)^2 - \frac{8\Phi_{11}}{c_2^3(\psi_2c_1)^2}\sigma_\epsilon^4\Delta_n,\\
&\widetilde{U}(j_1,j_2)_i^n = \widetilde{\eta}(j_1,1)_i^n\widetilde{\eta}(j_2,1)_i^n, \quad j_1 \neq j_2,\ j_1,j_2 = 1,\ldots,4,
\end{aligned}$$

where $\widetilde{\eta}(4,1)_i^n = \eta(4,1)_i^n$. For $j_1, j_2 = 1,\ldots,4$, $\widetilde{U}(j_1,j_2)_i^n$ approximate $U(j_1,j_2)_i^n$, which are defined as follows: for $a_q(p) \le i < a_{q+1}(p)$,

$$\begin{aligned}
&U(1,1)_i^n = (\eta(1,1)_i^n)^2 - \frac{2}{3}\widetilde{\sigma}^2_{a_q(p)\Delta_n}\Delta_n, \quad U(2,2)_i^n = (\eta(2,1)_i^n)^2 - \frac{8c_2\Phi_{22}}{(\psi_2c_1)^2}\sigma^4_{a_q(p)\Delta_n}\Delta_n,\\
&U(3,3)_i^n = (\eta(3,1)_i^n)^2 - \frac{16\Phi_{12}\sigma_\epsilon^2}{c_2(\psi_2c_1)^2}\sigma^2_{a_q(p)\Delta_n}\Delta_n, \quad U(4,4)_i^n = (\eta(4,1)_i^n)^2 - \frac{8\Phi_{11}}{c_2^3(\psi_2c_1)^2}\sigma_\epsilon^4\Delta_n,\\
&U(j_1,j_2)_i^n = \eta(j_1,1)_i^n\eta(j_2,1)_i^n, \quad j_1 \neq j_2,\ j_1,j_2 = 1,\ldots,4.
\end{aligned}$$

Note that inequalities (A.11) and (A.13) of Lemma 1 also hold when $\breve{\eta}$ is replaced by $\tilde{\eta}$ and K_r is replaced by a constant $K_{r,p}$; and that inequalities (A.12) also hold when η is replaced by $\tilde{\eta}$. Then we have the following lemma.

Lemma 2. *Suppose that Assumption 2 holds and $\theta_1 = 3/4$. Then, for any fixed p, we have*

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\left(U(j_1,j_2)_i^n - \widetilde{U}(j_1,j_2)_i^n\right) \xrightarrow{P} 0, \quad j_1,j_2 = 1,2,3, \tag{A.31}$$

as $n \to \infty$.

Proof. Step 1. We prove the case $j_1 = j_2 = 1$ of (A.31). Note that $a^2 - b^2 = 2b(a-b) + (a-b)^2$, then it is sufficient to show

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\left(\eta(1,1)_i^n - \widetilde{\eta}(1,1)_i^n\right)\widetilde{\eta}(1,1)_i^n \xrightarrow{P} 0, \tag{A.32}$$

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\left(\eta(1,1)_i^n - \widetilde{\eta}(1,1)_i^n\right)^2 \xrightarrow{P} 0. \tag{A.33}$$

By Lemma 1, it is not hard to obtain the result (A.33). Therefore, it is sufficient to show (A.32).

For $a_q(p) \le i < a_{q+1}(p)$, we have

$$\begin{aligned}\eta(1,1)_i^n - \widetilde{\eta}(1,1)_i^n &= \frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\widetilde{b}_u du ds + \frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}(\widetilde{\sigma}_u - \widetilde{\sigma}_{a_q(p)\Delta_n})d\widetilde{W}_u ds \\ &+ \frac{1}{\psi_2 k_n\Delta_n}\left[\sum_{j=l_n-k_n}^{l_n}\sum_{m=j+k_n-l_n}^{k_n-1}(g_m^n)^2(\Delta^n_{(i+1)l_n+j}C - \Delta^n_{il_n+j}C)\right] \\ &+ \frac{1}{\psi_2 k_n\Delta_n}\left[\sum_{j=1}^{k_n-1}\sum_{m=1}^{j}(g_m^n)^2(\Delta^n_{(i+1)l_n+j}C - \Delta^n_{il_n+j}C)\right] + O_P\left(\frac{k_n\sqrt{l_n\delta_n}}{l_n}\right) \\ &=: \alpha(1,1)_i^n + \alpha(1,2)_i^n + \alpha(1,3)_i^n + \alpha(1,4)_i^n + O_P\left(\frac{k_n\sqrt{l_n\delta_n}}{l_n}\right),\end{aligned}$$

where $\alpha(1,j)_i^n, j = 1,\ldots,4$ are the corresponding terms, respectively. For the argument in (A.32), Lemma 1 implies that it is sufficient to show that

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\alpha(1,j)_i^n\widetilde{\eta}(1,1)_i^n \xrightarrow{P} 0, \quad j = 1,2,3,4. \tag{A.34}$$

For the case $j = 1$ of (A.34), note that $\widetilde{b}_u = (\widetilde{b}_u - \widetilde{b}_{i\Delta_n}) + \widetilde{b}_{i\Delta_n}$. Then, it is sufficient to show that

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\widetilde{b}_{i\Delta_n} du ds \widetilde{\eta}(1,1)_i^n \xrightarrow{P} 0, \tag{A.35}$$

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}(\widetilde{b}_u - \widetilde{b}_{i\Delta_n}) du ds \widetilde{\eta}(1,1)_i^n \xrightarrow{P} 0. \tag{A.36}$$

For (A.35), we have

$$\begin{aligned}&E\left(n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\widetilde{b}_{i\Delta_n} du ds \widetilde{\eta}(1,1)_i^n\right)^2 \\ &= n^{\frac{1}{4}}\sum_{i=0}^{[t/\Delta_n]-2} EE^n_{il_n}\left(\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\widetilde{b}_{i\Delta_n} du ds \widetilde{\eta}(1,1)_i^n\right)^2 \\ &\le K n^{\frac{1}{4}} \times \frac{1}{\Delta_n} \times \Delta_n^2 \times \Delta_n = K n^{\frac{1}{4}}\Delta_n^2 \longrightarrow 0\end{aligned}$$

as n goes to infinity. For (A.36), two applications of Hölder's inequality suggest that

$$\begin{aligned}E^n_{il_n}\left(\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}(\widetilde{b}_u - \widetilde{b}_{i\Delta_n}) du ds\right)^2 &\le E^n_{il_n}\left(\int_0^{\Delta_n}\int_{i\Delta_n}^{(i+2)\Delta_n}|\widetilde{b}_u - \widetilde{b}_{i\Delta_n}| du ds\right)^2 \\ &\le K\Delta_n^4 E_i^n(\beta_i^n(p))^2,\end{aligned}$$

where

$$\beta_i^n(p) = \sup_{s,t\in[i\Delta_n,(i+2(p+2))\Delta_n]}\{|\widetilde{b}_s - \widetilde{b}_t| + |\sigma_s - \sigma_t| + |\widetilde{\sigma}_s - \widetilde{\sigma}_t| + |\overline{\sigma}_s - \overline{\sigma}_t|\}.$$

Then,

$$\begin{aligned}&E\left|n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}(\widetilde{b}_u - \widetilde{b}_{i\Delta_n}) du ds \widetilde{\eta}(1,1)_i^n\right| \\ &\le K n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\Delta_n\sqrt{E^n_{il_n}(\beta_i^n(p))^2} \times \Delta_n^{1/2} \le K\Delta_n\sum_{i=1}^{[t/2\Delta_n]}\sqrt{E^n_{il_n}(\beta_i^n(p))^2} \longrightarrow 0\end{aligned}$$

as $n \longrightarrow \infty$, where Lemma 5.4 of Jacod et al. (2009) is used in the last step. Hence, we have proved (A.34) for $j = 1$.

For $j = 2$ of (A.34), note that $d\widetilde{\sigma}_t = \overline{b}dt + \overline{\sigma}_t d\overline{W}_t$. Thus, it is sufficient to show that

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\int_{a_q(p)\Delta_n}^{u}\overline{b}_r dr dW_u ds \widetilde{\eta}(1,1)_i^n \xrightarrow{P} 0, \tag{A.37}$$

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\int_{a_q(p)\Delta_n}^{u}\overline{\sigma}_r d\overline{W}_r dW_u ds \widetilde{\eta}(1,1)_i^n \xrightarrow{P} 0. \tag{A.38}$$

For (A.37), the boundedness of $\bar{b}$ and Hölder's inequality imply that

$$E^n_{s_q}\left(\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\int_{a_q(p)\Delta_n}^{u}\bar{b}_r drdW_u ds\right)^2 \le K\Delta_n^3.$$

Further, by Lemma 1, we have

$$\begin{aligned}&En^{\frac{1}{8}}\left|\sum_{i=0}^{[t/\Delta_n]-2}\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\int_{a_q(p)\Delta_n}^{u}\bar{b}_r drdW_u ds\widetilde{\eta}(1,1)^n_i\right|\\&\le Kn^{\frac{1}{8}}\times\frac{1}{\Delta_n}\times\Delta_n^{3/2}\times\Delta_n^{1/2}=Kn^{\frac{1}{8}}\times\Delta_n\longrightarrow 0\end{aligned}$$

as n goes to infinity. Hence, we obtain (A.37). For (A.38), note that $\overline{\sigma}_r=(\overline{\sigma}_r-\overline{\sigma}_{a_q(p)\Delta_n})+\overline{\sigma}_{a_q(p)\Delta_n}$. Then, it is sufficient to show that

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\int_{a_q(p)\Delta_n}^{u}\overline{\sigma}_{a_q(p)\Delta_n}d\overline{W}_r dW_u ds\widetilde{\eta}(1,1)^n_i\xrightarrow{P}0, \tag{A.39}$$

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\int_{a_q(p)\Delta_n}^{u}(\overline{\sigma}_r-\overline{\sigma}_{a_q(p)\Delta_n})d\overline{W}_r dW_u ds\widetilde{\eta}(1,1)^n_i\xrightarrow{P}0. \tag{A.40}$$

We prove (A.39) first. Denote

$$\xi^n_i=\frac{1}{\Delta_n}\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\int_{a_q(p)\Delta_n}^{u}d\overline{W}_r dW_u ds.$$

By the property of Itô integral, we have $E^n_{s_q}\xi^n_i\widetilde{\eta}(1,1)^n_i=0$, and hence $E\xi^n_{i_1}\widetilde{\eta}(1,1)^n_{i_1}\xi^n_{i_2}\widetilde{\eta}(1,1)^n_{i_2}=0$ for $|i_1-i_2|>2(p+2)$. Therefore, similarly to (A.35), (A.39) follows from Hölder's inequality and Lemma 1. As to (A.40), by Hölder's inequality, we have

$$E^n_{s_q}\left(\int_0^{\Delta_n}\int_{i\Delta_n+s}^{(i+1)\Delta_n+s}\int_{a_q(p)\Delta_n}^{u}(\overline{\sigma}_r-\overline{\sigma}_{a_q(p)\Delta_n})d\overline{W}_r dW_u ds\right)^2\le K(\Delta_n)^4E^n_{s_q}(\beta^n_{a_q(p)}(p))^2.$$

Then, by method similar to the procedure used to prove (A.36), we obtain (A.40) by Hölder's inequality and applying Lemma 5.4 in Jacod et al. (2009). Hence, we have proved (A.34) for $j=2$.

For cases $j=3$ and $j=4$ of (A.34), we only prove the latter and the former follows similarly. The boundedness of σ implies that

$$|\alpha(1,4)^n_i|\le\frac{1}{\psi_2k_n\Delta_n}\sum_{j=1}^{k_n-1}\sum_{m=1}^{j}(g^n_m)^2\left|\int_0^{\delta_n}(\sigma^2_{s+((i+1)l_n+j)\delta_n}-\sigma^2_{s+(il_n+j)\delta_n})ds\right|\le\frac{K\beta^n_i(p)}{k_n\Delta_n}\le K\Delta_n\beta^n_i(p).$$

Then, by similar argument to the procedure used to prove (A.36), we obtain (A.34) for $j=4$.

Step 2. We prove the case $j_1=j_2=2$ in (A.31). Similar to Step 1, Lemma 1 implies that it is sufficient to show that

$$n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\left(\eta(2,1)^n_i-\widetilde{\eta}(2,1)^n_i\right)\widetilde{\eta}(2,1)^n_i\xrightarrow{P}0. \tag{A.41}$$

Denote $\bar{i}(j)=((i+1)l_n+j)\delta_n$ and $i(j)=(il_n+j)\delta_n$. For $a_q(p)\le i<a_{q+1}(p)$, set

$$\begin{aligned}\alpha(2,1)^n_i&=\frac{1}{\psi_2k_n\Delta_n}\sum_{j=0}^{l_n-k_n-1}[B(n,((i+1)l_n+j)\delta_n)_{k_n\delta_n}-B(n,(il_n+j)\delta_n)_{k_n\delta_n}],\\ \alpha(2,2)^n_i&=\frac{2}{\psi_2k_n\Delta_n}\sum_{j=0}^{l_n-k_n-1}\int_{\bar{i}(j)}^{\bar{i}(j)+k_n\delta_n}\int_{\bar{i}(j)}^{u}b_sg_n(s-\bar{i}(j))ds\sigma_ug_n(u-\bar{i}(j))dW_u,\\ \alpha(2,3)^n_i&=\frac{2}{\psi_2k_n\Delta_n}\sum_{j=0}^{l_n-k_n-1}\int_{\bar{i}(j)}^{\bar{i}(j)+k_n\delta_n}\int_{\bar{i}(j)}^{u}(\sigma_s\sigma_u-\sigma^2_{a_q(p)\Delta_n})g_n(s-\bar{i}(j))dW_sg_n(u-\bar{i}(j))dW_u,\end{aligned}$$

and let $\alpha(2,4)^n_i$ and $\alpha(2,5)^n_i$ be the same formulae as $\alpha(2,2)^n_i$ and $\alpha(2,3)^n_i$ when $\bar{i}(j)$ is replaced by $i(j)$, respectively. Then, we have

$$\eta(2,1)^n_i-\widetilde{\eta}(2,1)^n_i=\alpha(2,1)^n_i+\alpha(2,2)^n_i+\alpha(2,3)^n_i-\alpha(2,4)^n_i-\alpha(2,5)^n_i.$$

For (A.41), it is sufficient to show that

$$n^{\frac{1}{8}} \sum_{i=0}^{[t/\Delta_n]-2} \alpha(2,j)_i^n \widetilde{\eta}(2,1)_i^n \xrightarrow{P} 0, \quad j=1,\ldots,5. \tag{A.42}$$

For the case $j=1$ in (A.42), we have

$$\begin{aligned}
&E_{il_n}^n(\alpha(2,1)_i^n)^2 \\
&\le \frac{K}{(k_n\Delta_n)^2} \sum_{j_1,j_2=0}^{l_n-k_n-1} E_{il_n}^n B(n,(il_n+j_1)\delta_n)_{k_n\delta_n} B(n,(il_n+j_2)\delta_n)_{k_n\delta_n} \\
&\le \frac{Kk_n}{(k_n\Delta_n)^2} \sum_{j=0}^{l_n-k_n-1} E_{il_n}^n \left(\int_{i(j)}^{i(j)+k_n\delta_n} \int_{i(j)}^{u} \sigma_s g_n(s-i(j)) dW_s b_u g_n(u-i(j)) du \right)^2 + O_P(\delta_n) \\
&\le \frac{Kk_n l_n}{(k_n\Delta_n)^2} (k_n\delta_n)^3 + O_P(\delta_n) \le K n^{-3/4}.
\end{aligned}$$

Then, Lemma 1 implies that (A.42) holds for $j=1$.

Similar to the case where $j=1$, we can obtain (A.42) for $j=2$ and $j=4$.

For cases $j=3$ and $j=5$ in (A.42), we only prove the latter, as the former can be proved in a similar manner. Let

$$\begin{aligned}
\alpha(2,6)_i^n &= \frac{2\sigma_{a_q(p)\Delta_n}}{\psi_2 k_n\Delta_n} \sum_{j=0}^{l_n-k_n-1} \int_{i(j)}^{i(j)+k_n\delta_n} \int_{i(j)}^{u} g_n(s-i(j)) dW_s (\sigma_u - \sigma_{a_q(p)\Delta_n}) g_n(u-i(j)) dW_u, \\
\alpha(2,7)_i^n &= \frac{2}{\psi_2 k_n\Delta_n} \sum_{j=0}^{l_n-k_n-1} \int_{i(j)}^{i(j)+k_n\delta_n} \int_{i(j)}^{u} (\sigma_s - \sigma_{a_q(p)\Delta_n}) g_n(s-i(j)) dW_s g_n(u-i(j)) \sigma_u dW_u.
\end{aligned}$$

Then, $\alpha(2,5)_i^n = \alpha(2,6)_i^n + \alpha(2,7)_i^n$, and it is sufficient to prove (A.42) for cases $j=6$ and $j=7$.

We now prove the case of $j=6$ in (A.42). Note that under Assumption 2, by Itô's formula, there exist two locally bounded càdlàg processes $\underline{b}, \underline{\sigma}$ such that $d\sigma_t = \underline{b}_t dt + \underline{\sigma}_t d\widetilde{W}_t$; see, for example, Aït-Sahalia and Jacod (2014). Set

$$\begin{aligned}
\alpha(2,8)_i^n &= \frac{2\sigma_{a_q(p)\Delta_n}}{\psi_2 k_n\Delta_n} \sum_{j=0}^{l_n-k_n-1} \int_{i(j)}^{i(j)+k_n\delta_n} \int_{i(j)}^{u} g_n(s-i(j)) dW_s \int_{a_q(p)\Delta_n}^{u} \underline{b}_r dr g_n(u-i(j)) dW_u, \\
\alpha(2,9)_i^n &= \frac{2\sigma_{a_q(p)\Delta_n}}{\psi_2 k_n\Delta_n} \sum_{j=0}^{l_n-k_n-1} \int_{i(j)}^{i(j)+k_n\delta_n} \int_{i(j)}^{u} g_n(s-i(j)) dW_s \int_{a_q(p)\Delta_n}^{u} \underline{\sigma}_r d\widetilde{W}_r g_n(u-i(j)) dW_u.
\end{aligned}$$

Then, $\alpha(2,6)_i^n = \alpha(2,8)_i^n + \alpha(2,9)_i^n$, and it is sufficient to show (A.42) for $j=8$ and $j=9$. For $j=8$, we have

$$\begin{aligned}
&E_{s_q}^n(\alpha(2,8)_i^n)^2 \\
&\le \frac{Kk_n\sigma^2_{a_q(p)\Delta_n}}{(k_n\Delta_n)^2} \sum_{j=0}^{l_n-k_n-1} E_{s_q}^n \left(\int_{i(j)}^{i(j)+k_n\delta_n} \int_{i(j)}^{u} g_n(s-i(j)) dW_s \int_{a_q(p)\Delta_n}^{u} \underline{b}_r dr g_n(u-i(j)) dW_u \right)^2 \\
&\le K\delta_n.
\end{aligned}$$

Hence, Lemma 1 implies that (A.42) holds for $j=8$. For $j=9$, note that $\underline{\sigma}_r = \underline{\sigma}_{a_q(p)\Delta_n} + (\underline{\sigma}_r - \underline{\sigma}_{a_q(p)\Delta_n})$, and denote $\alpha(2,10)_i^n$ and $\alpha(2,11)_i^n$ by the term $\alpha(2,9)_i^n$, where $\int_{i\Delta_n}^{u} \underline{\sigma}_r d\widetilde{W}_r$ is replaced by $\int_{i\Delta_n}^{u} \underline{\sigma}_{a_q(p)\Delta_n} d\widetilde{W}_r$ and $\int_{i\Delta_n}^{u} (\underline{\sigma}_r - \underline{\sigma}_{a_q(p)\Delta_n}) d\widetilde{W}_r$, respectively. Then, $\alpha(2,9)_i^n = \alpha(2,10)_i^n + \alpha(2,11)_i^n$. For the case where $j=10$ in (A.42), similar to the proof of (A.35), by Lemma 1 and the following result:

$$\begin{aligned}
&E_{s_q}^n(\alpha(2,10)_i^n)^2 \\
&\le \frac{Kk_n}{(k_n\Delta_n)^2} \sum_{j=0}^{l_n-k_n-1} E_{s_q}^n \left(\int_{i(j)}^{i(j)+k_n\delta_n} \int_{i(j)}^{u} g_n(s-i(j)) dW_s \int_{a_q(p)\Delta_n}^{u} \underline{\sigma}_{a_q(p)\Delta_n} d\widetilde{W}_r g_n(u-i(j)) dW_u \right)^2 \\
&\le \frac{Kk_n l_n}{(k_n\Delta_n)^2} \times (k_n\delta_n)^2 (l_n\delta_n) \le Kk_n\delta_n,
\end{aligned}$$

we obtain (A.42) for $j = 10$. For $j = 11$ in (A.42), similar to proving (A.36), the desired result is not hard to obtain. Hence, we have proved (A.42) for $j = 6$.

For $j = 7$ in (A.42), similar to the procedure for $j = 6$, we can show that the argument holds. Hence, we have proved (A.42) for $j = 5$, obtained (A.41), and completed the proof of step 2.

Step 3. We prove the case $j_1 = j_2 = 3$ in (A.31). Similar to Step 1, Lemma 1 implies that it is sufficient to show that

$$n^{\frac{1}{8}} \sum_{i=0}^{[t/\Delta_n]-2} \left(\eta(3,1)_i^n - \widetilde{\eta}(3,1)_i^n\right) \widetilde{\eta}(3,1)_i^n \xrightarrow{P} 0.$$

For $a_q(p) \le i < a_{q+1}(p)$, set

$$\alpha(3,1)_i^n = \frac{2}{\psi_2 k_n \Delta_n} \sum_{j=0}^{l_n-k_n-1} [\overline{X^c}_{i(j)}^n - \sigma_{a_q(p)\Delta_n} \overline{W}_{i(j)}^n] \overline{\epsilon}_{i(j)}^n,$$

$$\alpha(3,2)_i^n = \frac{2}{\psi_2 k_n \Delta_n} \sum_{j=0}^{l_n-k_n-1} [\overline{X^c}_{i(j)}^n - \sigma_{a_q(p)\Delta_n} \overline{W}_{i(j)}^n] \overline{\epsilon}_{i(j)}^n,$$

and we have $\eta(3,1)_i^n - \widetilde{\eta}(3,1)_i^n = \alpha(3,1)_i^n - \alpha(3,2)_i^n$. Then, it is sufficient to show that

$$n^{\frac{1}{8}} \sum_{i=0}^{[t/\Delta_n]-2} \alpha(3,j)_i^n \widetilde{\eta}(3,1)_i^n \xrightarrow{P} 0, \quad j = 1, 2. \tag{A.43}$$

As the two cases are similar, we only prove the case where $j = 2$ in (A.43). Because

$$E_{il_n}^n \left(\frac{2}{\psi_2 k_n \Delta_n} \sum_{j=0}^{l_n-k_n-1} \int_{i(j)}^{i(j)+k_n\delta_n} b_u g_n(u - i(j)) du \overline{\epsilon}_{i(j)}^n \right)^2 \le \frac{K}{l_n},$$

Lemma 1 implies that it is sufficient to show that (A.43) holds for $\alpha(3,3)_i^n$ with

$$\alpha(3,3)_i^n = \frac{2}{\psi_2 k_n \Delta_n} \sum_{j=0}^{l_n-k_n-1} \int_{i(j)}^{i(j)+k_n\delta_n} (\sigma_u - \sigma_{a_q(p)\Delta_n}) g_n(u - i(j)) dW_u \overline{\epsilon}_{i(j)}^n.$$

Then, similar to the proof of (A.42) for $j = 6$, it is not hard to show that (A.43) holds for $j = 3$. Hence, we have proved (A.31) for $j_1 = j_2 = 3$.

Step 4. For the case $j_1 \neq j_2$ in (A.31), similar to the above procedure, it is sufficient to show that

$$n^{\frac{1}{8}} \sum_{i=0}^{[t/\Delta_n]-2} \left(\eta(j_1,1)_i^n - \widetilde{\eta}(j_1,1)_i^n\right) \widetilde{\eta}(j_2,1)_i^n \xrightarrow{P} 0, \quad j_1 \neq j_2.$$

It is not difficult to obtain the above results through the procedure of Steps 1–3. □

We now prove Theorem 1.

Proof of Theorem 1. Note that, from Proposition 2, the consistency of the estimator is obvious. As to the CLT result (12), it is sufficient to show that

$$U_t^n =: n^{1/8} \left(V(\nu;\theta_1,1)_t^n - \frac{2}{c_1^2} \int_0^t \gamma_s^2 ds - \frac{2}{3} \int_0^t \widetilde{\sigma}_s^2 ds \right) \xrightarrow{L-s} \int_0^t \Sigma_s dB_s, \tag{A.44}$$

$$n^{\frac{1}{8}} \left(\widehat{\int_0^t \gamma_s^2 ds} - \int_0^t \gamma_s^2 ds \right) \xrightarrow{P} 0, \tag{A.45}$$

as $n \to \infty$. The result (A.45) is implied by

$$n^{\frac{1}{4}} \left(\widehat{\int_0^t \gamma_s^2 ds} - \int_0^t \gamma_s^2 ds \right) = O_P(1), \tag{A.46}$$

which follows easily from Theorem 3 of Podolskij and Vetter (2009a). Then, it remains to show (A.44). Recalling that $\Delta_n = l_n\delta_n$, we have

$$\begin{aligned} U_t^n &= n^{\frac{1}{8}}\left(\sum_{i=0}^{[t/\Delta_n]-2}(\eta(1,1)_i^n)^2 - \frac{2}{3}\int_0^t \widetilde{\sigma}_s^2 ds\right) + n^{\frac{1}{8}}\left(\sum_{i=0}^{[t/\Delta_n]-2}(\eta(2,1)_i^n)^2 - \frac{8c_2\Phi_{22}}{(\psi_2 c_1)^2}\int_0^t \sigma_s^4 ds\right) \\ &\quad + n^{\frac{1}{8}}\left(\sum_{i=0}^{[t/\Delta_n]-2}(\eta(3,1)_i^n)^2 - \frac{16\Phi_{12}\sigma_\epsilon^2}{c_2(\psi_2 c_1)^2}\int_0^t \sigma_s^2 ds\right) + n^{\frac{1}{8}}\left(\sum_{i=0}^{[t/\Delta_n]-2}(\eta(4,1)_i^n)^2 - \frac{8\Phi_{11}\sigma_\epsilon^4 t}{c_2^3(\psi_2 c_1)^2}\right) \\ &\quad + \sum_{j_1\neq j_2=1}^{4} n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\eta(j_1,1)_i^n\eta(j_2,1)_i^n \\ &\quad + n^{\frac{1}{8}}\sum_{i=0}^{[t/\Delta_n]-2}\sum_{j_1=5}^{8}\eta(j_1,1)_i^n\left(2\sum_{j_2=1}^{4}\eta(j_2,1)_i^n + \sum_{j_1=5}^{8}\eta(j_1,1)_i^n\right) \\ &=: \sum_{j=1}^{4}U(j,j)_t^n + \sum_{j_1\neq j_2=1}^{4}U(j_1,j_2)_t^n + R_{1,t}^n, \end{aligned} \tag{A.47}$$

where $U(j_1,j_2)_t^n, j_1, j_2 = 1, 2, 3, 4$, and $R_{1,t}^n$ are the corresponding terms, respectively. Note that $U(j_1,j_2)_t^n = U(j_2,j_1)_t^n$.

Recall the notations right before Lemma 2. $\widetilde{U}(j_1,j_2)_t^n$ is an approximation of $U(j_1,j_2)_t^n$, where $\eta(j,1)_i^n$ is replaced by $\widetilde{\eta}(j,1)_i^n$. Then, we can write

$$\widetilde{U}(j_1,j_2)_t^n = n^{\frac{1}{8}}\sum_{q=1}^{J_n(p)}\sum_{i=a_q(p)}^{b_q(p)-1}\widetilde{U}(j_1,j_2)_i^n + n^{\frac{1}{8}}\sum_{q=1}^{J_n(p)}\sum_{i=b_q(p)}^{a_{q+1}(p)-1}\widetilde{U}(j_1,j_2)_i^n + n^{\frac{1}{8}}\widetilde{R}(j_1,j_2)_t^n,$$

where, for $j_1, j_2 = 1, 2, 3, 4$,

$$\begin{aligned} \widetilde{R}(1,1)_t^n &= \sum_{i=J_n(p)}^{[t/\Delta_n]-2}(\widetilde{\eta}(1,1)_i^n)^2 + \frac{2}{3}\sum_{q=1}^{J_n(p)}\widetilde{\sigma}_{a_q(p)\Delta_n}^2 2(p+2)\Delta_n - \frac{2}{3}\int_0^t \widetilde{\sigma}_s^2 ds, \\ \widetilde{R}(2,2)_t^n &= \sum_{i=J_n(p)}^{[t/\Delta_n]-2}(\widetilde{\eta}(2,1)_i^n)^2 + \frac{8c_2\Phi_{22}}{(\psi_2 c_1)^2}\sum_{q=1}^{J_n(p)}\sigma_{a_q(p)\Delta_n}^4 2(p+2)\Delta_n - \frac{8c_2\Phi_{22}}{(\psi_2 c_1)^2}\int_0^t \sigma_s^4 ds, \\ \widetilde{R}(3,3)_t^n &= \sum_{i=J_n(p)}^{[t/\Delta_n]-2}(\widetilde{\eta}(3,1)_i^n)^2 + \frac{16\Phi_{12}\sigma_\epsilon^2}{c_2(\psi_2 c_1)^2}\sum_{q=1}^{J_n(p)}\sigma_{a_q(p)\Delta_n}^2 2(p+2)\Delta_n - \frac{16\Phi_{12}\sigma_\epsilon^2}{c_2(\psi_2 c_1)^2}\int_0^t \sigma_s^2 ds, \\ \widetilde{R}(4,4)_t^n &= \sum_{i=J_n(p)}^{[t/\Delta_n]-2}(\widetilde{\eta}(4,1)_i^n)^2 + \frac{8\Phi_{11}}{c_2^3(\psi_2 c_1)^2}\sigma_\epsilon^4\sum_{q=1}^{J_n(p)}2(p+2)\Delta_n - \frac{8\Phi_{11}}{c_2^3(\psi_2 c_1)^2}\sigma_\epsilon^4 t, \\ \widetilde{R}(j_1,j_2)_t^n &= \sum_{i=J_n(p)}^{[t/\Delta_n]-2}\widetilde{\eta}(j_1,1)_i^n\widetilde{\eta}(j_2,1)_i^n, \quad j_1\neq j_2,\ j_1,j_2 = 1,\ldots,4. \end{aligned}$$

It is not difficult to show that $n^{\frac{1}{8}}\widetilde{R}(j_1,j_2)_t^n$ converges to zero in probability as n and p go to infinity, for $j_1, j_2 = 1, \ldots, 4$. Note that the first part of $\widetilde{U}(j_1,j_2)_t^n$ is the dominant term, and that the dominant terms $\sum_{i=a_q(p)}^{b_q(p)-1}\widetilde{U}(j_1,j_2)_i^n$ are uncorrelated for different q. To prove result (A.44), it is sufficient to show that

$$U(j_1,j_2)_t^n - \widetilde{U}(j_1,j_2)_t^n \xrightarrow{P} 0, \quad j_1,j_2 = 1,\ldots,4, \tag{A.48}$$

$$n^{\frac{1}{4}}\sum_{q=1}^{J_n(p)}E_{s_q}^n\left(\sum_{i=a_q(p)}^{b_q(p)-1}\left(\sum_{j=1}^{4}\widetilde{U}(j,j)_i^n + 2\sum_{j_1,j_2=1,j_1<j_2}^{4}\widetilde{U}(j_1,j_2)_i^n\right)\right)^2 \xrightarrow{P} \int_0^t \Sigma_s^2 ds, \tag{A.49}$$

$$n^{\frac{1}{2}}\sum_{q=1}^{J_n(p)}E_{s_q}^n\left(\sum_{i=a_q(p)}^{b_q(p)-1}\left(\sum_{j=1}^{4}\widetilde{U}(j,j)_i^n + 2\sum_{j_1,j_2=1,j_1<j_2}^{4}\widetilde{U}(j_1,j_2)_i^n\right)\right)^4 \xrightarrow{P} 0, \tag{A.50}$$

$$n^{\frac{1}{8}}\sum_{q=1}^{J_n(p)}E_{s_q}^n\sum_{i=a_q(p)}^{b_q(p)-1}\left(\sum_{j=1}^{4}\widetilde{U}(j,j)_i^n + 2\sum_{j_1,j_2=1,j_1<j_2}^{4}\widetilde{U}(j_1,j_2)_i^n\right)(N_{b_q(p)\Delta_n} - N_{a_q(p)\Delta_n}) \xrightarrow{P} 0, \tag{A.51}$$

$$n^{\frac{1}{8}} \sum_{q=1}^{J_n(p)} \sum_{i=b_q(p)}^{a_{q+1}(p)-1} \widetilde{U}(j_1, j_2)_i^n \xrightarrow{P} 0, \quad j_1, j_2 = 1, \dots, 4, \tag{A.52}$$

$$n^{\frac{1}{8}} \sum_{i=0}^{[t/\Delta_n]-2} \eta(j_1, 1)_i^n \eta(j_2, 1)_i^n \xrightarrow{P} 0, \quad j_1 = 5, \dots, 8, j_2 = 1, \dots, 8, \tag{A.53}$$

as n and p go to infinity, where N is any component of $(W, \widetilde{W})$ or a bounded martingale orthogonal to W and $\widetilde{W}$. Lemma 2 implies (A.48).

Step 1. For (A.49), we can show that, for $j_1, \dots, j_4 = 1, \dots, 4$,

$$n^{\frac{1}{4}} \sum_{q=1}^{J_n(p)} E_{s_q}^n \left(\sum_{i=a_q(p)}^{b_q(p)-1} \widetilde{U}(j_1, j_2)_i^n \sum_{i=a_q(p)}^{b_q(p)-1} \widetilde{U}(j_3, j_4)_i^n \right) \xrightarrow{P} \int_0^t \Sigma(j_1, j_2, j_3, j_4)_s^2 ds, \tag{A.54}$$

as n and p go to infinity, where $\Sigma(j_1, j_2, j_3, j_4)_t = \Sigma(j_2, j_1, j_3, j_4)_t$, $\Sigma(j_1, j_2, j_3, j_4)_t = \Sigma(j_1, j_2, j_4, j_3)_t$, and

$$\begin{aligned}
&\Sigma(1,1,1,1)_t^2 = c_1 \widetilde{\sigma}_t^4, \quad \Sigma(2,2,2,2)_t^2 = \frac{192(\Phi_{22} c_2)^2}{\psi_2^4 c_1^3} \sigma_t^8, \\
&\Sigma(3,3,3,3)_t^2 = \frac{768 \Phi_{12}^2 \sigma_\epsilon^4}{\psi_2^4 c_1^3 c_2^2} \sigma_t^4, \quad \Sigma(4,4,4,4)_t^2 = \frac{192 \Phi_{11}^2}{\psi_2^4 c_1^3 c_2^6} \sigma_\epsilon^8, \\
&\Sigma(1,2,1,2)_t^2 = \frac{16 \Phi_{22} c_2}{\psi_2^2 c_1} \sigma_t^4 \widetilde{\sigma}_t^2, \quad \Sigma(1,3,1,3)_t^2 = \frac{32 \Phi_{12} \sigma_\epsilon^2}{\psi_2^2 c_1 c_2} \sigma_t^2 \widetilde{\sigma}_t^2, \\
&\Sigma(1,4,1,4)_t^2 = \frac{16 \Phi_{11} \sigma_\epsilon^4}{\psi_2^2 c_2^3 c_1} \widetilde{\sigma}_t^2, \quad \Sigma(2,3,2,3)_t^2 = \frac{192 \Phi_{12} \Phi_{22} \sigma_\epsilon^2}{\psi_2^4 c_1^3} \sigma_t^6, \\
&\Sigma(2,4,2,4)_t^2 = \frac{96 \Phi_{11} \Phi_{22} \sigma_\epsilon^4}{c_2^2 \psi_2^4 c_1^3} \sigma_t^4, \quad \Sigma(3,4,3,4)_t^2 = \frac{192 \Phi_{11} \Phi_{12} \sigma_\epsilon^6}{c_2^4 \psi_2^4 c_1^3} \sigma_t^2, \\
&\Sigma(j_1, j_2, j_3, j_4)_t^2 = 0, \quad (j_1, j_2) \neq (j_3, j_4).
\end{aligned}$$

After some standard and complicated computations, it is not hard to prove that (A.54) holds. By (A.47) and the expression of γ_t^2, we derive $\Sigma_t^2 = \sum_{j_1,\dots,j_4=1}^4 \Sigma(j_1, j_2, j_3, j_4)_t^2$ and obtain (A.49).

Step 2. For (A.50), using the C_r inequality, it is sufficient to show that

$$n^{\frac{1}{2}} \sum_{q=1}^{J_n(p)} E_{s_q}^n \left(\sum_{i=a_q(p)}^{b_q(p)-1} \widetilde{U}(j_1, j_2)_i^n \right)^4 \xrightarrow{P} 0, \quad j_1, j_2 = 1, \dots, 4, \tag{A.55}$$

as n goes to infinity. For the case $j_1 = j_2 = 1$, by Lemma 1, we have

$$E_i^n (U(j_1, j_2)_i^n)^4 \le K(l_n \delta_n)^4.$$

Hence, we obtain (A.55) for the case $j_1 = j_2 = 1$ because $b_q(p) - a_q(p) = 2p$ is independent of n. For the other cases, we can similarly use Lemma 1 to obtain the desired results.

Step 3. For (A.51), it is sufficient to show that

$$n^{\frac{1}{8}} \sum_{q=1}^{J_n(p)} E_{s_q}^n \sum_{i=a_q(p)}^{b_q(p)-1} \widetilde{U}(j_1, j_2)_i^n (N_{b_q(p)\Delta_n} - N_{a_q(p)\Delta_n}) \xrightarrow{P} 0, \quad j_1, j_2 = 1, \dots, 4. \tag{A.56}$$

Similar to the procedure used to prove (6.25) in Vetter (2015), it is not hard to obtain (A.56) by Itô's lemma and the property of the normal distribution.

Step 4. For (A.52), it is sufficient to show that

$$n^{\frac{1}{4}} \sum_{q=1}^{J_n(p)} E_{s_q}^n \left(\sum_{i=b_q(p)}^{a_{q+1}(p)-1} \widetilde{U}(j_1, j_2)_i^n \right)^2 \xrightarrow{P} 0, \quad j_1, j_2 = 1, \dots, 4,$$

as n and p go to infinity. Similar to Step 1, we obtain the above results as n and p go to infinity.

Step 5. For (A.53), Hölder's inequality suggests that it is sufficient to show that

$$n^{\frac{1}{8}} \sum_{i=0}^{[t/\Delta_n]-2} (\eta(j, 1)_i^n)^2 \xrightarrow{P} 0, \quad j = 5, \dots, 8. \tag{A.57}$$

For $j = 5$ in (A.57), Lemma 1 implies that (A.57) holds for $j = 5$. For $j = 6$, similar to the proof of the case $j = 6$ in (A.21), the condition $\theta_2 < 1/4$ implies that the result holds. For $j = 7$, similar to the case $j = 7$ in (A.21), the condition $\theta_2 > 3(1-\theta_1)/(4(2-\beta))$ suggests that (A.57) holds for $j = 7$. By Hölder's inequality, we obtain result (A.57) for $j = 8$ from the results for $j = 6$ and $j = 7$. □

Remark 12. By very similar arguments to that of the proof for Theorem 1, we can show that the following general results hold under Assumption 2 and $\widetilde{\sigma} \equiv 0$. That is,

(a) if $1/2 < \theta_1 < 4/5$ and $\theta_1/(4(2-\beta)) < \theta_2 < 1/4$, then for any fixed t,

$$n^{\frac{1-\theta_1}{2}}\left(n^{2\theta_1-\frac{3}{2}}V(\nu;\theta_1,m)_t^n - \frac{2}{c_1^2}\int_0^t \gamma_s^2 ds\right) \xrightarrow{L-s} \sqrt{\frac{12}{c_1^3}}\int_0^t \gamma_s^2 dB_s;$$

(b) if $\theta_1 = 4/5$ and $\theta_1/(4(2-\beta)) < \theta_2 < 1/4$, then for any fixed t,

$$n^{\frac{1-\theta_1}{2}}\left(n^{2\theta_1-\frac{3}{2}}V(\nu;\theta_1,m)_t^n - \frac{2}{c_1^2}\int_0^t \gamma_s^2 ds\right) \xrightarrow{L-s} m^2 c_1\int_0^t \widetilde{b}_u^2 du + \sqrt{\frac{12}{c_1^3}}\int_0^t \gamma_s^2 dB_s; \tag{A.58}$$

(c) if $4/5 < \theta_1 < 1$ and $(1-\theta_1)/(2-\beta) < \theta_2 < 1/4$, then for any fixed t,

$$n^{\frac{5}{2}-3\theta_1}\left(n^{2\theta_1-\frac{3}{2}}V(\nu;\theta_1,m)_t^n - \frac{2}{c_1^2}\int_0^t \gamma_s^2 ds\right) \xrightarrow{P} m^2 c_1\int_0^t \widetilde{b}_u^2 du.$$

In fact, for the proof of the above results, one only needs to make straightforward adjustments for the following quantities in the proof of Theorem 1 and hence the related subsequent calculations: $J_n(p) = [([t/\Delta_n] - m - 2)/(2(p+m+1))]$, $a_q(p) = 2(q-1)(p+m+1)$, $b_q(p) = a_q(p) + 2p$ (see (A.30)). Note that the terms $\eta(j,m)_i^n$ for $j = 5,\ldots,8$ are always negligible in proving the asymptotic theory of $V(\nu;\theta_1,m)_t^n$. In the case of $\widetilde{\sigma} \equiv 0$, for $1/2 < \theta_1 < 4/5$, by Lemma 1, one can easily find that the three terms $\eta(j,m)_i^n, j = 2,3,4$ have the same magnitude and are the dominant terms; for $4/5 < \theta_1 < 1$, $\eta(1,m)_i^n$ is the dominant term; for $\theta_1 = 4/5$ the four terms $\eta(j,m)_i^n, j = 1,2,3,4$ have the same magnitude and are the dominant terms.

Proof of Theorem 2. The theorem follows straightforwardly from the stable convergence in Theorem 1 and result (15). □

Proof of Proposition 3. The proposition follows straightforwardly from (A.58) and result (A.46). The derivation is quite similar to the proof of Theorem 1 and hence omitted. □

Proof of Theorem 3. The theorem follows straightforwardly from Proposition 3 and (22). □

Appendix B. Asymptotic property of the test statistic T_n when volatility is rough

In this section, we study the asymptotic property of the test statistic T_n under rough volatility alternatives. We consider the following rough volatility specification for the efficient log price process

$$\begin{cases} dX_t = b_t\, dt + \sigma_t\, dW_t + dJ_t, \\ d\ln\sigma_t = \widetilde{b}_t\, dt + \widetilde{\sigma}_t\, dB_t^H, \quad 0 \le t \le T, \end{cases} \tag{B.1}$$

where B_t^H is a fractional Brownian motion with Hurst parameter H, and the integral $\int_0^t \widetilde{\sigma}_s\, dB_s^H$ should be interpreted as a pathwise Riemann–Stieltjes integral. We assume that $\widetilde{\sigma}$ is a process with paths of finite $\underline{q}$-variation with $\underline{q} < 1/(1-H)$, then $\int_0^t \sigma_s dB_s^H$ is well defined as a pathwise Riemann–Stieltjes integral (see, e.g., Corcuera et al., 2006, Liu and Jing, 2018). When $0 < H < 1/2$, the paths of B^H are rough, yielding rough volatility processes.

Assumption 3. The observed log price process and the latent log price process are given by (2) and (B.1) with $0 < H < 1/2$, respectively. The conditions (ii) and (iii) of Assumption 1 hold. The drift processes b and $\widetilde{b}$ and the volatility processes σ and $\widetilde{\sigma}$ are all adapted, locally bounded, and càdlàg. $\sigma_t > 0$ and $\widetilde{\sigma}_t > 0$ for all $t \ge 0$. $\widetilde{\sigma}$ is a continuous process with paths of finite $\underline{q}$-variation with $\underline{q} < 1/(1-H)$.

We have the following theorem.

Theorem 4. *Suppose that Assumption 3 holds, $\theta_1 > (4H+1)/(4H+2)$ and $(1-\theta_1)H/(2-\beta) < \theta_2 < 1/4$. Then,*

$$T_n \asymp_P n^{\frac{2}{5}(1-H)} \quad \text{and} \quad T_n \xrightarrow{P} \infty,$$

where $a_n \asymp_P b_n$ means $a_n/b_n = O_P(1)$ and $b_n/a_n = O_P(1)$.

Proof. By (B.1) and the continuity of $\widetilde{\sigma}$, for $t > s$, we have

$$\begin{aligned}\sigma_t^2 - \sigma_s^2 &= \exp(\ln\sigma_s^2)(\exp(\ln\sigma_t^2 - \ln\sigma_s^2) - 1) = \sigma_s^2(\ln\sigma_t^2 - \ln\sigma_s^2 + O_P((\ln\sigma_t^2 - \ln\sigma_s^2)^2)) \\ &= 2\sigma_s^2(\int_s^t \widetilde{b}_u du + \int_s^t \widetilde{\sigma}_u dB_u^H) + O_P((t-s)^{2H}) \\ &= 2\sigma_s^2\widetilde{\sigma}_s(B_t^H - B_s^H) + o_P((t-s)^H) + O_P((t-s)) + O_P((t-s)^{2H}).\end{aligned}$$

Using this result and recalling (A.14), we have

$$\begin{aligned}\eta(1,m)_i^n &= \frac{1}{\Delta_n}\int_0^{\Delta_n}(\sigma^2_{(i+m)\Delta_n+s} - \sigma^2_{i\Delta_n+s})ds + O_P\left(\frac{1}{k_n\Delta_n}\right) \\ &= 2\sigma^2_{i\Delta_n}\widetilde{\sigma}_{i\Delta_n}\frac{1}{\Delta_n}\int_0^{\Delta_n}(B^H_{(i+m)\Delta_n+s} - B^H_{i\Delta_n+s})ds + o_P(\Delta_n^H) + O_P\left(\frac{1}{k_n\Delta_n}\right).\end{aligned} \tag{B.2}$$

The result (B.2) plus (A.10), (A.12) and (A.27) leads to the following

$$\widehat{\nu}_{(i+1)\Delta_n} - \widehat{\nu}_{i\Delta_n} = 2\sigma^2_{i\Delta_n}\widetilde{\sigma}_{i\Delta_n}\frac{1}{\Delta_n}\int_0^{\Delta_n}(B^H_{(i+1)\Delta_n+s} - B^H_{i\Delta_n+s})ds + o_P(\Delta_n^H) + O_P(n^{-\theta_2(2-s)}).$$

Note that by the self-similarity property of B^H, $B^H_{(i+m+s)\Delta_n} - B^H_{(i+s)\Delta_n}$ has the same distribution as $\Delta_n^H(B^H_{i+m+s} - B^H_{i+s})$. Therefore,

$$\frac{1}{\Delta_n}\int_0^{\Delta_n}(B^H_{(i+m)\Delta_n+s} - B^H_{i\Delta_n+s})ds \overset{d}{=} \Delta_n^H\int_0^1(B^H_{i+m+s} - B^H_{i+s})ds.$$

Then by using similar arguments of the proof of Theorem 1 in Corcuera et al. (2006), we obtain that when $\theta_1 > (4H+1)/(4H+2)$ and $(1-\theta_1)H/(2-\beta) < \theta_2 < 1/4$,

$$n^{(\theta_1-1)(1-2H)}V(\nu;\theta_1,m)_t^n \xrightarrow{P} 4c_1^{2H-1}\overline{\nu}_{m,H}\int_0^t \sigma_s^4\widetilde{\sigma}_s^2 ds, \tag{B.3}$$

where $\overline{\nu}_{m,H} = E(\int_0^1(B^H_{i+m+s} - B^H_{i+s})ds)^2$. Let $d_H = 4c_1^{2H-1}\overline{\nu}_{m,H}\int_0^t\sigma_s^4\widetilde{\sigma}_s^2ds$, hence, we have $V(\nu;\theta_1,m)_t^n = (d_H + o_P(1)) \times n^{(1-\theta_1)(1-2H)}$. Similarly, we have $PV(\nu,m)_t^n = O_P(n^{1/2-H})$ and $BV(\nu)_t^n = O_P(n^{1/2-H})$ for $\theta_1 = 3/4$.

Finally, by (B.3), for $H < 1/2$, we have

$$\begin{aligned}&n^{1/5}V(\nu;4/5,1)_t^n = n^{1/5}\times(d_H + o_P(1))\times n^{(1-\frac{4}{5})(1-2H)} = (d_H + o_P(1))\times n^{\frac{2}{5}(1-H)}, \\ &\widehat{\mathcal{B}}_t = n^{1/6}\times O_P(n^{(1-\frac{5}{6})(1-2H)}) = O_P\left(n^{\frac{1}{3}(1-H)}\right), \\ &\widehat{\Upsilon}_t = \min\left\{\frac{27}{c_1^3}\widehat{\int_0^t\gamma_s^4ds}, 9BV(\nu)_t^n/2\right\} \xrightarrow{P} \frac{27}{c_1^3}\int_0^t\gamma_s^4ds, \\ &\widehat{\int_0^t\gamma_s^2ds} \xrightarrow{P} \int_0^t\gamma_s^2ds,\end{aligned}$$

which lead to

$$T_n = \frac{n^{1/5}\frac{3}{2}\left(V(\nu;4/5,1)_t^n - \frac{2n^{-1/10}}{c_1^2}\widehat{\int_0^t\gamma_s^2ds}\right) - c_1\frac{3}{2}\widehat{\mathcal{B}}_t}{\sqrt{\widehat{\Upsilon}_t}} = \frac{\frac{3}{2}d_H + o_P(1)}{\sqrt{\frac{27}{c_1^3}\int_0^t\gamma_s^4ds}}\times n^{\frac{2}{5}(1-H)} + O_P\left(n^{\frac{1}{3}(1-H)}\right),$$

completing the proof. □

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Rev. Financial Stud. 9 (2), 385–426.
Aït-Sahalia, Y., Fan, J., Li, Y., 2013. The leverage effect puzzle: Disentangling sources of bias at high frequency. J. Financ. Econom. 109, 224–249.
Aït-Sahalia, Y., Fan, J., Peng, H., 2009. Nonparametric transition-based tests for jump diffusions. J. Amer. Statist. Assoc. 104 (487), 1102–1116.
Aït-Sahalia, Y., Jacod, J., 2014. High-Frequency Financial Econometrics. Princeton University Press.
Aït-Sahalia, Y., Park, J.Y., 2012. Stationarity-based specification tests for diffusions when the process is nonstationary. J. Econometrics 169 (2), 279–292.
Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Internat. Econom. Rev. 39 (4), 885–905.
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 (6), 1481–1536.
Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (2), 253–280.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econom. 2 (1), 1–37.
Barndorff-Nielsen, O.E., Veraart, A., 2009. Stochastic Volatility of Volatility in Continuous Time. CREATES research paper.
Bibinger, M., Winkelmann, L., 2018. Common price and volatility jumps in noisy high-frequency data. Electron. J. Stat. 12 (1), 2018–2073.
Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. J. Political Econ. 81 (3), 637–654.
Bull, A.D., 2017. Semimartingale detection and goodness-of-fit tests. Ann. Statist. 45 (3), 1254–1283.
Chen, Q., Hu, M., Song, X., 2019. A nonparametric specification test for the volatility functions of diffusion processes. Econometric Rev. 38 (5), 557–576.
Christensen, K., Hounyo, U., Podolskij, M., 2018. Is the diurnal pattern sufficient to explain the intraday variation in volatility: A nonparametric assessment. J. Econometrics 205, 336–362.
Clinet, S., Potiron, Y., 2020. Estimation for high-frequency data under parametric market microstructure noise. Ann. Inst. Statist. Math. forthcoming.
Cont, R., Kokholm, T., 2013. A consistent pricing model for index options and volatility derivatives. Math. Finance 23 (2), 248–274.
Corcuera, J.M., Nualart, D., Woerner, J.H.C., 2006. Power variation of some integral fractional processes. Bernoulli 12 (4), 713–735.
Corradi, V., Swanson, N.R., 2005. Bootstrap specification tests for diffusion processes. J. Econometrics 124 (1), 117–148.
Corradi, V., White, H., 1999. Specification tests for the variance of a diffusion. J. Time Series Anal. 20 (3), 253–270.
Corsi, F., Pirino, D., Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. J. Econometrics 159 (2), 276–288.
Dette, H., Podolskij, M., 2008. Testing the parametric form of the volatility in continuous time diffusion models–a stochastic process approach. J. Econometrics 143 (1), 56–73.
Dette, H., Podolskij, M., Vetter, M., 2006. Estimation of integrated volatility in continuous-time financial models with applications to goodness-of-fit testing. Scand. J. Stat. 33 (2), 259–278.
Dette, H., von Lieres und Wilkau, C., 2003. On a test for a parametric form of volatility in continuous time financial models. Finance Stoch. 7 (3), 363–384.
Fan, J., Wang, Y., 2008. Spot volatility estimation for high-frequency data. Stat. Interface 1 (2), 279–288.
Foster, D.P., Nelson, D.B., 1996. Continuous record asymptotics for rolling sample variance estimators. Econometrica 64 (1), 139–174.
Gatheral, J., Jaisson, T., Rosenbaum, M., 2018. Volatility is rough. Quant. Finance 18 (6), 933–949.
Gloter, A., Jacod, J., 2001. Diffusions with measurement errors. II: Optimal estimators. ESAIM Probab. Stat. 5, 243–260.
Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. 6 (2), 327–343.
Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Rev. Financ. Stud. 18 (1), 37–84.
Jacod, J., Li, P.A., Mykland, M., Podolskij, Y., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. Stochastic Process. Appl. 119, 2249–2276.
Jacod, J., Li, Y., Zheng, X., 2017. Statistical properties of microstructure noise. Econometrica 85, 1133–1174.
Jacod, J., Li, Y., Zheng, X., 2019. Estimating the integrated volatility with tick observations. J. Econometrics 208 (1), 80–100.
Jacod, J., Protter, P., 2012. Discretization of Processes. Springer Berlin Heidelberg.
Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes. Springer-Verlag, Berlin.
Jing, B.Y., Liu, Z., Kong, X.B., 2014. On the estimation of integrated volatility with jumps and microstructure noise. J. Bus. Econom. Statist. 32 (3), 457–467.
Li, Y., Xie, S., Zheng, X., 2016. Efficient estimation of integrated volatility incorporating trading information. J. Econometrics 195 (1), 33–50.
Li, Y., Zhang, Z., Li, Y., 2018. A unified approach to volatility estimation in the presence of both rounding and random market microstructure noise. J. Econometrics 203 (2), 187–222.
Liu, G., Jing, B.Y., 2018. On estimation of Hurst parameter under noisy observations. J. Bus. Econom. Statist. 36, 483–492.
Liu, Q., Liu, Y., Liu, Z., 2017. Estimating spot volatility in the presence of infinite variation jumps. Stochastic Process. Appl. 128 (6), 1958–1987.
Mancini, C., 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scand. J. Stat. 36 (2), 270–296.
Mancini, C., Mattiussi, V., Renò, R., 2015. Spot volatility estimation using delta sequences. Finance Stoch. 19 (2), 261–293.
Mancino, M.E., Recchioni, M.C., Sanfelici, S., 2017. Fourier–Malliavin Volatility Estimation: Theory and Practice. Springer.
Munk, A., Schmidt-Hieber, J., 2010. Lower bounds for volatility estimation in microstructure noise models. In: Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown 6. Institute of Mathematical Statistics, pp. 43–55.
Mykland, P.A., Zhang, L., 2017. Assessment of uncertainty in high frequency data: The observed asymptotic variance. Econometrica 85 (1), 197–231.
Papanicolaou, A., Giesecke, K., 2016. Variation-based tests for volatility misspecification. J. Econometrics 191 (1), 217–230.
Podolskij, M., Vetter, M., 2009a. Bipower-type estimation in a noisy diffusion setting. Stochastic Process. Appl. 119 (9), 2803–2831.
Podolskij, M., Vetter, M., 2009b. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli 15 (3), 634–658.
Sanfelici, S., Curato, I.V., Mancino, M.E., 2015. High-frequency volatility of volatility estimation free from spot volatility estimates. Quant. Finance 15 (8), 1–15.
Song, Z., 2011. A martingale approach for testing diffusion models based on infinitesimal operator. J. Econometrics 162 (2), 189–212.
Vasicek, O., 1977. An equilibrium characterization of the term structure. J. Financ. Econom. 5 (2), 177–188.
Vetter, M., 2015. Estimation of integrated volatility of volatility with applications to goodness-of-fit testing. Bernoulli 21 (4), 2393–2418.
Vetter, M., Dette, H., 2012. Model checks for the volatility under microstructure noise. Bernoulli 18 (4), 1421–1447.
Wang, C.D., Mykland, P.A., 2014. The estimation of leverage effect with high-frequency data. J. Amer. Statist. Assoc. 109 (505), 197–215.
Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159, 235–250.
Zhang, L., 2011. Estimating covariation: Epps effect, microstructure noise. J. Econometrics 160 (1), 33–47.
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 (472), 1394–1411.
Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. J. Bus. Econom. Statist. 14 (1), 45–52.
Zu, Y., Boswijk, H.P., 2014. Estimating spot volatility with high-frequency financial data. J. Econometrics 181 (2), 117–135.