Journal of Econometrics 208 (2019) 80–100

Contents lists available at ScienceDirect

Journal of Econometrics

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

Estimating the integrated volatility with tick observations

Jean Jacod [a], Yingying Li [b], Xinghua Zheng [b,*]

[a] CNRS – UMR 7586, and Université Pierre et Marie Curie - P6, France
[b] Hong Kong University of Science and Technology, Hong Kong

ARTICLE INFO

Article history:
Available online 5 October 2018

JEL classification:
C14
C13
D40

Keywords:
High frequency data
Integrated volatility
Market microstructure noise
Dependent noise
Endogenous time

ABSTRACT

We develop a volatility estimator that can be directly applied to tick-by- tick data. More specifically, we consider a model that allows for (i) irregular observation times that can be endogenous, (ii) dependent noise that can have diurnal features and be dependent on the latent price process, and (iii) jumps in the latent price process. We show that our estimator yields consistent estimates and enjoys the optimal rate of convergence. Simulation as well as empirical studies demonstrate favorable properties of our proposed estimator.

1. Introduction

The estimation of the integrated volatility based on ultra high frequency data has attracted tremendous attention recently and great progress has been made.

In the first papers devoted to the subject, the noise was supposed to be an additive white noise, independent of the latent process; the observation times were assumed to be equidistant or slightly more generally, irregular but independent of the latent price process which was assumed to contain no jump. One can quote, among others, the two-scale realized volatility in Zhang et al. (2005), the multi-scale realized volatility in Zhang (2006), the realized kernel estimators in Barndorff-Nielsen et al. (2008), or the quasi-maximum likelihood estimator introduced in Xiu (2010); the latter three classes of estimators achieve the optimal convergence rate $n_s^{1/4}$, where n_s is the number of observed returns used in the estimation. The pre-averaging method of Jacod et al. (2009) also features the optimal rate $n_s^{1/4}$, under weaker assumptions on the noise, which basically require that conditionally on the latent process the noise is centered and independent, although not necessarily identically distributed (so, unconditionally, the noise may be dependent, and may also depend on the latent process).

We remark that it may not be appropriate to apply these pioneer methods directly to tick-by-tick data. Oftentimes, in order to satisfy the conditions assumed, data cleaning procedures that involve substantial subsampling are necessary, such as reducing tick-by-tick data with sample size n to minute-by-minute or at best second-by-second data with sample size $n_s \ll n$. As an example, for Citigroup Inc (NYSE:C) over the month of January 2011, the average size of complete tick-by-tick data is about $n = 246{,}000$ per day, compared with second-by-second data where $n_s = 23{,}400$, we have $n_s \approx n^{4/5}$. This means if tick-by-tick data can be directly used, even a sub-optimal rate of convergence $n^{1/5}$ would result in the same efficiency as the optimal rate achieved by second-by-second data $n_s^{1/4}$. The method we present in this paper would do even better, achieving the optimal rate of convergence with tick-by-tick data, i.e., $n^{1/4}$.

* Correspondence to: Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.
E-mail address: xhzheng@ust.hk (X. Zheng).

https://doi.org/10.1016/j.jeconom.2018.09.006
0304-4076/

There are two distinctive features that tick data have: (i) they are irregularly spaced and the observation times may depend on price process, and more importantly (ii) the market microstructure noise may be dependent (see Hansen and Lunde (2006) and Jacod et al. (2017)). Of course, dependency for the noise may occur for regularly spaced observations, but for minute-by-minute data for example it is much less prominent than for tick-by-tick data. In addition, (iii) jumps exist in the latent price process; see, e.g., Aït-Sahalia et al. (2012).

Developments have been made to relax the assumptions. Among other works, Li et al. (2014), Li et al. (2013) studied endogenous observation times, Kalnina and Linton (2008) and Barndorff-Nielsen et al. (2011) considered a specific form of dependent noise, and Jing et al. (2014) incorporated jumps with noisy observations. Li et al. (2016) considered a situation where the noise can be (partially) modeled as a function of trading information while allowing for dependent noise, endogenous observation times and jumps. Li et al. (2018) studied the case when both rounding and random market microstructure noise are present.

There is however lack of work that gives an ultimate estimator which can be used directly in the case when the following three features are present simultaneously: (i) observation times can be irregular and even endogenous, (ii) market microstructure noise can be dependent, and (iii) the latent price process can have jumps. In particular, at the optimal convergence rate $n^{1/4}$, the noise induces a bias term which needs to be removed and which strongly depends on the autocovariance structure of the noise.

Our aim in this work is to develop a volatility estimator that works in the presence of the above three features (i), (ii) and (iii).

Let us discuss a bit more about the dependence of microstructure noise. In Jacod et al. (2017), we studied the estimation of the (joint) moments of microstructure noise. The empirical studies therein revealed (moderate) positive autocorrelations of the microstructure noise for the stocks that were tested, in particular, the studies show that the noise is not white but colored. Earlier works in this direction include Hansen and Lunde (2006), Ukabata and Oya (2009) and Aït-Sahalia et al. (2011). The simplest model that accommodates dependent noise is again an additive noise independent of the latent process, but with a non-trivial autocovariance structure. In such a setting, assuming an exponential decay of the noise autocovariance, Aït-Sahalia et al. (2011) extend the two-scale estimator, with the rate $n_s^{1/6}$, and the multi-scale estimator, with the optimal rate $n_s^{1/4}$ (but without providing a "feasible" limit theorem). Barndorff-Nielsen et al. (2011) consider a multivariate setting and substantially generalize the model for noise in Kalnina and Linton (2008). As they study the multivariate setting in which case asynchronicity arises, their model further allows for irregular observation times. To ensure that the final estimator is positive (semi-definite in the multivariate case), a wider bandwidth is adopted, which leads to a suboptimal rate. Hautsch and Podolskij (2013) consider the case when the noise is stationary and q-dependent with some known value q, extends the pre-averaging estimator and shows that it achieves $n_s^{1/4}$ convergence rate. In another paper devoted to the multivariate case, Christensen et al. (2013) combine the pre-averaging idea with the Hayashi–Yoshida estimator and obtain a rate-optimal covariance matrix estimator. In Remark 3.3 therein the authors explain that their estimator is still consistent if the noise is finite-dependent. The authors however did not pursue Central Limit Theorem (CLT) in this case. Ikeda (2015, 2016) proposes two-scale realized kernels for both the univariate and multivariate cases. The noise is assumed to take a product form similar to (2.7). He also allows for irregular observations, however with a strong condition which restricts the applicability.[1]

Our assumptions on the noise are significantly weaker than the papers quoted above. Furthermore, we relax the assumption that the noise is independent of the latent process. The dependence is modeled via a multiplicative semimartingale process, which can be used to capture, for example, diurnal patterns of the noise. In particular, the noise in our model does not need to be stationary. About observation times, we allow them to be endogenous and, in particular, they can be irregularly spaced. We develop a volatility estimator that can overcome these challenges and hence can be applied directly to tick-by-tick data. We further prove the consistency and an associated feasible central limit theorem with the optimal rate $n^{1/4}$.

As we will see, in order to achieve the optimal rate of convergence we need to incorporate some information on the autocovariance function of the noise into the estimator. This has of course to be estimated from the data, and it is done by using the autocovariance function estimators which have been proposed in our earlier paper Jacod et al. (2017). It is worth pointing out that Jacod et al. (2017) also allow for irregular observation times and jumps. Below, we freely use the results of that paper.

We also conduct both simulation and empirical studies to examine the performance of our proposed volatility estimator. Comparison is made between our new estimator with the traditional pre-averaging estimator. We pick pre-averaging estimator as a representative pioneer method because it is rate-optimal and has been shown in various applications to perform well and be robust. From our comparisons, it is seen clearly that when the noise is colored, our proposed estimator is nearly unbiased whereas the traditional pre-averaging estimator can have a significant bias. We have also intentionally added rounding effect, which violates the assumptions of this estimator. It is seen that with this additional rounding effect, the proposed estimator still performs well. Furthermore, empirical studies clearly show an important advantage of our estimator, which is that it yields stable estimates when applied to different sampling frequencies even when directly applied to tick-by-tick data, whereas the traditional pre-averaging estimator may yield rather different estimates.

[1] In Assumption (A-4') in Ikeda (2015) as well as Assumption (A-2)(a) in Ikeda (2016), with $\Delta t_i = t_i - t_{i-1}$ and $N = \max\{i : t_i \leq 1\}$, the author assumes that Δt_i can be written as D_i/N and D_i is adapted. The latter assumption implies that $N \cdot \Delta t_i$ is adapted. We find this condition difficult to interpret since N is observed only at time 1.

The paper is organized as follows. The setting is described in Section 2. Our proposed estimator, which is a modified pre-averaged estimator, is introduced in Section 3. Section 3 also contains the main results. Simulation and empirical studies are given in Sections 4 and 5 respectively. Section 6 concludes. The scheme of the proofs is given in Appendix A, whereas the details are provided in a supplement available on the web.

2. Setting and assumptions

We have three basic ingredients: the underlying process X, typically the log-price of an asset, the observation scheme, and the noise. All random quantities below are defined on a fixed filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

Before stating the assumption, we recall the general form of a possibly discontinuous Itô semimartingale Y on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Such a process admits a Grigelionis decomposition of the form

$$
\begin{aligned}
Y_t = Y_0 + \textstyle\int_0^t b_s^Y\, ds + \int_0^t \sigma_s^Y\, dW_s^Y + \int_{[0,t]\times E} \delta^Y(s,z)\, 1_{\{|\delta^Y(s,z)|\leq 1\}}\, (\underline{p} - \underline{q})(ds, dz) \\
+ \textstyle\int_{[0,t]\times E} \delta^Y(s,z)\, 1_{\{|\delta^Y(s,z)|> 1\}}\, \underline{p}(ds, dz).
\end{aligned} \tag{2.1}
$$

Here, W^Y is a standard Brownian motion, $\underline{p}$ is a Poisson random measure on $\mathbb{R}_+ \times E$, where $(E, \mathcal{E})$ is a Polish space (i.e., a complete separable metric space, one could always take $E = \mathbb{R}$, or even $E = [0, 1]$), with a non-random intensity measure of the form $\underline{q}(dt, dz) = dt \otimes \lambda(dz)$ with λ a σ-finite measure on $(E, \mathcal{E})$. The processes b^Y and σ^Y are optional; see for example Jacod and Shiryaev (2003) for this notion: if those processes are càdlàg adapted, they are a fortiori optional. The function δ^Y on $\Omega \times \mathbb{R}_+ \times E$ is predictable, and they are such that all integrals in (2.1) make sense.

The following assumption on a process Y of the form (2.1) depends on a parameter $r \in [0, 2]$, which sets a bound on the degree of activity of jumps (when $r = 2$ there are essentially no restrictions on the jumps, when $r \leq 1$ they are summable on each finite interval, when $r = 0$ they have finite activity).

Assumption (K-r) on Y: There are a sequence τ_m of stopping times increasing to ∞, a sequence of *deterministic* nonnegative λ-integrable functions J_m^Y on E, and a sequence w_m of real numbers, such that $|\, b_t^Y(\omega)| \leq w_m$ and $|\,\sigma_t^Y(\omega)| \leq w_m$ and $|\,\delta^Y(\omega, t, z)|^r \wedge 1 \leq J_m^Y(z)$ for all (ω, t, z) with $t \leq \tau_m(\omega)$. □

This assumption will be used for a number of different processes below. Note that if we consider several processes Y at once, their driving Brownian motions W^Y are usually different, possibly correlated, but it is always possible to find a single Poisson measure $\underline{p}$ which drives the jumps of all Y's, hence the notation $\underline{p}$ instead of $\underline{p}^Y$.

In particular, the underlying process X is of the form (2.1), and will satisfy (K-r) for some $r \in [0, 1)$, so in that formula we can dispense with the integral with respect with $\underline{p} - \underline{q}$ (up to modifying the drift term) and it can be written as follows (for simplicity we write its coefficients and driving Brownian motion as $b,\ \sigma,\ \delta$ and W:

$$
X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s + \int_{[0,t]\times E} \delta(s,z)\underline{p}(ds, dz). \tag{2.2}
$$

We need the following assumption, with some $r \in [0, 1)$:

Assumption (H-r): The process X is of the form (2.2), satisfies (K-r), and both b and σ satisfy (K-2). □

The rather unusual, but reasonably weak, requirement that the drift b is an Itô semimartingale is used for dealing with irregular observation times; it could be weakened, but we need a nice behavior of $\mathbb{E}(b_T - b_S \mid \mathcal{F}_S)$ and $\mathbb{E}((b_T - b_S)^2 \mid \mathcal{F}_S)$ in terms of $\mathbb{E}(T - S \mid \mathcal{F}_S)$ for finite stopping times $S \leq T$, which is complicated to state in general but satisfied under (K-2).

Next, we describe how observations take place. At stage n, that is, for a given frequency of observations, the successive observations occur at times $0 = T(n, 0) < T(n, 1) < \cdots$ for a sequence $T(n, i)$ of (possibly random) finite times, and we set

$$
N_t^n = \sum_{i\geq 1} 1_{\{T(n,i)\leq t\}}, \qquad \Delta(n, i) = T(n, i) - T(n, i - i). \tag{2.3}
$$

That is, the number of observations up to time t is $N_t^n + 1$, and $\Delta(n, i)$ is the ith inter-observation lag. So far, this is simply notation, and we will make the hypothesis that, at stage n, the time lags $\Delta(n, i)$ are, in an appropriate sense, of the same order of magnitude Δ_n for all i, where Δ_n is a *fixed* sequence of positive numbers going to 0. Note, however, that the $T(n, i)$'s and $\Delta(n, i)$'s are observed, at least up to the time horizon t, whereas Δ_n is a *non-observable* mathematical abstraction, which should not enter the various statistics used by the observer.

The precise assumption, which involves two numbers $\rho,\ \rho'$, goes as follows:

Assumption (O-$\rho,\ \rho'$): All $T(n, i)$'s are stopping times, and there is an Itô semimartingale (α_t) satisfying (K-2), with $\alpha_t > 0$ and $\alpha_{t-} > 0$ for all $t > 0$ and such that

(i) $\Delta_n N_t^n \xrightarrow{\mathbb{P}} A_t := \int_0^t \alpha_s\, ds$ for all t.

(ii) For all $s, t > 0$, the sequence $\Delta_n^{1/2+\rho'}\big(N_t^n - N^n_{(t-s\Delta_n^{\rho'})+}\big)$ is bounded in probability.

(iii) For any $\kappa \geq 2$, there are a sequence $(\tau(\kappa)_m)_{m\geq 1}$ of stopping times increasing to ∞ and real numbers $(w(\kappa)_m)$ such that we have for all i, n, m:

$$
T(n, i-1) \leq \tau(\kappa)_m \Rightarrow \begin{cases} \big|\,\mathbb{E}\big(\alpha_{T(n,i-1)}\Delta(n, i) \mid \mathcal{F}_{T(n,i-1)}\big) - \Delta_n\big| \leq w(\kappa)_m \Delta_n^{1+\rho} \\ \mathbb{E}\big(|\alpha_{T(n,i-1)}\Delta(n, i)|^\kappa \mid \mathcal{F}_{T(n,i-1)}\big) \leq w(\kappa)_m \Delta_n^{\kappa}. \end{cases} \qquad \square \tag{2.4}
$$

Assumptions (ii) and (iii) become weaker when ρ' increases and ρ decreases, but below we will require $\rho' < 1/2$ and $\rho > 1/4$. Note that (i) and (ii) are not compatible if $\rho' < 1/4$, so we implicitly assume $\rho' \geq 1/4$. Also Assumption (ii) with $\rho' = 1/4$ is satisfied in standard settings; see below for more discussions.

The requirements in (O-ρ, ρ') ensure that the sampling, up to the modulating process α_t and in a kind of statistical sense, is close enough to a regular sampling. Of special mention is the first part of (2.4), which ensures a minimal orthogonality between $\Delta(n, i)$ and the past before time $T(n, i-1)$, and turns out to be essential to get the desired stable convergence in law below.

This assumption is slightly stronger than Assumption (O) in Jacod et al. (2017). Of course, it accommodates *regular sampling schemes* and *Poisson schemes* (the $T(n, i)$'s are the jump times of a Poisson process with intensity $1/\Delta_n$), both with $\alpha_t \equiv 1$ and any $\rho > 0$ and $\rho' = 1/4$. It also accommodates, with $\rho' = 1/4$ in all cases, the following schemes: *modulated Poisson schemes* (with a Poisson process with stochastic intensity $\frac{1}{\Delta_n}\alpha_t$), with $\rho = 1$; *modulated random walk schemes* (when $\Delta(n, i+1) = \Delta_n \Phi_{i+1}/\alpha_{T(n,i)}$ and the Φ_i's are i.i.d. positive independent of everything else and with $\mathbb{E}(\Phi_i) = 1$ and finite moments of all orders), with any $\rho > 0$; *time-changed regular sampling schemes* (with $T(n, i) = \inf(t : A_t = i\Delta_n)$ and A_t as in (O)), with $\rho = 1$.

Finally, we turn to the noise. At time $T(n, i)$ the process X is contaminated by some noise, meaning that we observe the variable

$$Y_i^n = X_{T(n,i)} + \varepsilon_i^n, \tag{2.5}$$

where the noise is ε_i^n.

Before stating our precise assumptions, we recall the ρ-mixing property of a stationary sequence $\chi = (\chi_i)_{i\in\mathbb{Z}}$ of variables, indexed by $\mathbb{Z}$: letting $\mathcal{G}_j = \sigma(\chi_i : i \leq j)$ and $\mathcal{G}^j = \sigma(\chi_i : i \geq j)$ be the pre- and post-σ-fields at time j, the ρ-mixing coefficients of χ for $k \geq 1$ are

$$\begin{aligned}\rho_k(\chi) = \sup\big\{|\mathbb{E}(UV)| : \quad & \mathbb{E}(U) = \mathbb{E}(V) = 0,\ \mathbb{E}(U^2) \leq 1,\ \mathbb{E}(V^2) \leq 1, \\ & U \text{ is } \mathcal{G}_0\text{-measurable, } V \text{ is } \mathcal{G}^k\text{-measurable}\big\},\end{aligned} \tag{2.6}$$

and we say that χ is v-polynomially ρ-mixing if for some $K > 0$, $\rho_k(\chi) \leq K/k^v$ for all $k \geq 1$, where v is a positive number, which we will assume larger than 3.

Assumption (N-v): The noise $(\varepsilon_i^n)_{i\geq 0}$ can be realized as

$$\varepsilon_i^n \ = \ \gamma_{T(n,i)} \cdot \chi_i, \tag{2.7}$$

where γ is a nonnegative Itô semimartingale satisfying Assumption (K-2). Furthermore, $(\chi_i)_{i\in\mathbb{Z}}$ is a stationary process, independent of the σ-field $\mathcal{F}_\infty = \bigvee_{t>0} \mathcal{F}_t$, and with mean 0 and variance 1, with finite moments of all orders, and which is v-polynomially ρ-mixing.

The process γ is included to accommodate possible diurnal features of the noise so that the size of the noise may change over time, and also a size which may depend on the level of the log-price X_t, while the autocorrelation remains roughly unchanged. Such features are consistent with the empirical findings in our earlier paper Jacod et al. (2017). A similar model has been adopted in Zheng and Li (2011) and Xia and Zheng (2018) to capture diurnal patterns in volatility.

The convention $\mathbb{E}(\chi_i^2) = 1$ is made to resolve the non-identifiability built in the formulation (2.7), since one could multiply $\gamma_{T(n,i)}$ and divide χ_i by the same constant without modifying the variables ε_i^n. It is thus not a restriction.

Note that the noise can be *dependent* on X, a form of dependency being induced by the presence of the process γ. However, despite this dependency, the noise and the process X are uncorrelated, since $\mathbb{E}(\varepsilon_i^n f(X)) = 0$ for any function $f(X)$ of the whole path of X: this is a drawback of the model used here. See Remark 2.8 of Jacod et al. (2017) for more comments on Assumption (N-v).

We denote by $r(m) = \mathbb{E}(\chi_0\chi_m)$ the autocovariance of the sequence χ_i, so $r(0) = 1$ and $r(m) = r(-m)$. The ρ-mixing property classically implies for some constant K:

$$|r(m)| \leq \frac{K}{(|m|+1)^v}, \qquad \text{so } R = \sum_{m\in\mathbb{Z}} r(m) \text{ is well defined if } v > 1. \tag{2.8}$$

3. Pre-averaging and estimation of the integrated volatility

Our aim is to estimate the integrated volatility over some time interval $[0, t]$, that is the variable

$$C_t \ = \ \int_0^t \sigma_s^2\, ds.$$

For this, we use the pre-averaging method developed in Jacod et al. (2009), plus appropriate estimators for the autocovariance of the noise.

In order to implement the pre-averaging estimators, we need a weight (or, kernel) function g on $\mathbb{R}$, which satisfies

$$\begin{aligned}& g \text{ is continuous, piecewise } C^1 \text{ with a piecewise Lipschitz derivative } g', \\ & s \notin (0,1) \ \Rightarrow\ g(s) = 0, \qquad \textstyle\int_0^1 g(s)^2 ds \ > \ 0.\end{aligned}$$

We also need a sequence $h_n \geq 2$ of integers going to infinity, and we consider the following numbers (indexed by $n \geq 1$ and $i, j \in \mathbb{Z}$), and functions:

$$\begin{array}{lll} g_i^n = g(i/h_n), & \overline{g}_i^n \; = \; g_{i+1}^n - g_i^n & \\ \phi_j^n = \frac{1}{h_n} \sum_{i\in\mathbb{Z}} g_i^n \, g_{i-j}^n, & \overline{\phi}_j^n = h_n \sum_{i\in\mathbb{Z}} \overline{g}_i^n \, \overline{g}_{i-j}^n & \\ \phi(s) = \int g(u) g(u-s)\, du, & \overline{\phi}(s) = \int g'(u) g'(u-s)\, du & \\ \Phi_{00} = \int_0^1 \phi(s)^2 \, ds, & \Phi_{01} = \int_0^1 \phi(s)\overline{\phi}(s)\, ds, & \Phi_{11} = \int_0^1 \overline{\phi}(s)^2 \, ds. \end{array} \tag{3.1}$$

Both ϕ and $\overline{\phi}$ are even functions with support in $[-1, 1]$. We have, uniformly in $s \in \mathbb{R}$:

$$\phi_{[sh_n]}^n = \phi(s) + \mathrm{O}(1/h_n), \qquad \overline{\phi}_{[sh_n]}^n = \overline{\phi}(s) + \mathrm{O}(1/h_n). \tag{3.2}$$

In the simple case when $g(x) = x \wedge (1-x)$ for $x \in (0, 1)$, we have

$$\phi(0) = \frac{1}{12}, \quad \overline{\phi}(0) = 1, \quad \Phi_{00} = \frac{151}{80640}, \quad \Phi_{01} = \frac{1}{96}, \quad \text{and} \quad \Phi_{11} = \frac{1}{6},$$

see Eqs. (3.11) and (3.12) in Jacod et al. (2009).

Recall that we observe Y_i^n, as given by (2.5) and (2.7). More generally, for any process V we write $V_i^n = V_{T(n,i)}$, and also $\Delta_i^n V = V_i^n - V_{i-1}^n$ (so, for example, $\Delta_i^n X$ is the ith return, and $\Delta_i^n Y = Y_i^n - Y_{i-1}^n$ is the ith observed (noisy) return). If V_i^n is any array of variables, the pre-averaged values (implicitly depending on g) at stage n are

$$\widetilde{V}(h_n)_i^n = \sum_{j=1}^{h_n-1} g_j^n \, \Delta_{i+j}^n V = -\sum_{j=0}^{h_n-1} \overline{g}_j^n \, V_{i+j}^n. \tag{3.3}$$

A natural estimator for C_t is then the (properly normalized and also truncated in case of jumps) sum of all squares $(\widetilde{V}(h_n)_i^n)^2$. However, if we want to achieve the optimal rate, these estimators exhibit a bias due to the noise, as they already do in the case of regularly spaced observations, no jumps and a white noise. When the noise is colored, the de-biasing term is more involved. For this we use the following simple averages, based on another sequence $k_n \geq 2$ of integers increasing to infinity:

$$\overline{V}_i^n = \frac{1}{k_n} \sum_{j=0}^{k_n-1} V_{i+j}^n, \tag{3.4}$$

and for $m \in \{0, \ldots, k_n\}$ we set

$$U(m)_t^n = \sum_{i=0}^{N_t^n - 5k_n} (Y_i^n - \overline{Y}_{i+2k_n}^n)(Y_{i+m}^n - \overline{Y}_{i+4k_n}^n) \tag{3.5}$$

Actually, $U(m)$ is a proxy for $\sum_{i=0}^{N_t^n-5k_n} \gamma_i^n \chi_i \chi_{i+m}$, which itself is, after proper normalization, an estimator for $r(m) \int_0^t \gamma_s \, ds$: indeed, since γ and X do not vary "too much" in the time interval $[T(n,i), T(n, i+5k_n)]$, the ith summand in (3.5) is close to $\gamma_i^n(\chi_i^n - \overline{\chi}_{i+2k_n}^n)(\chi_{i+m}^n - \overline{\chi}_{i+4k_n}^n)$; since $m \leq k_n$, the two variables $\overline{\chi}_{i+2k_n}^n$ and $\overline{\chi}_{i+4k_n}^n$ are sufficiently close to 0 and also uncorrelated themselves and with χ_i^n and χ_{i+m}^n for the ith summand to be nearly $\gamma_i^n \chi_i^n \chi_{i+m}^n$.

As a matter of fact, we need pre-averaging for estimating C_t, and also for estimating the asymptotic variance, with two different window sizes $\overline{h}_n$ and $\overline{h}'_n$. We need the sequence k_n in (3.4) for de-biasing, and another sequence $k'_n \geq 2$ of integers going to infinity for the asymptotic variance again. The precise behavior of those sequences is as follows, for some real number $\theta > 0$:

$$\overline{h}_n \sim \frac{\theta}{\sqrt{\Delta_n}}, \qquad \overline{h}'_n \asymp \frac{1}{\Delta_n^\tau} \quad \text{with} \quad \frac{1}{2} < \tau \leq \frac{3}{5}, \qquad k_n \asymp \frac{1}{\Delta_n^{1/5}}, \qquad k'_n \asymp \frac{1}{\Delta_n^{1/8}}. \tag{3.6}$$

Here, $a_n \sim b_n$ is the usual equivalence and $a_n \asymp b_n$ means $1/K \leq a_n/b_n \leq K$ for some constant $K \geq 1$. In practice, Δ_n is not observable, and a good proxy for its order of magnitude is $1/N_1^n$.

Since our underlying process X may jump, in order to eliminate those jumps we use the truncation method of Mancini (2001) on the pre-averaged values. This is done by using yet two other tuning parameters $u_n, u'_n > 0$ satisfying

$$u_n \asymp (\overline{h}_n \Delta_n)^{\varpi}, \quad u'_n \asymp (\overline{h}'_n \Delta_n)^{\varpi} \qquad \text{with} \quad \frac{1}{4-2r} < \varpi < \frac{2[v]-3}{4[v]-4}. \tag{3.7}$$

Here, $[v]$ denotes the integer part of v and r is the bound on jumps activity in (H-r). If we know that X is continuous, then we do not need to truncate and we may take $u_n = u'_n = \infty$, although using the truncation above is innocuous in the continuous case. Otherwise, (3.7) requires $r < \frac{2[v]-4}{2[v]-3}$, which is smaller than 1 but goes to 1 as v increases to infinity: this is to be compared to the bound $r < 1$ for the same method when there is no noise.

We are now ready to exhibit our estimators for C_t, with ϕ_0^n and $\overline{\phi}_m^n$ associated with $h_n = \overline{h}_n$ in (3.1):

$$\widehat{C}_t^n = \frac{1}{\overline{h}_n\,\phi_0^n} \sum_{i=0}^{N_t^n-\overline{h}_n} (\widetilde{Y}(\overline{h}_n)_i^n)^2\, 1_{\{|\widetilde{Y}(\overline{h}_n)_i^n|\le u_n\}} - \frac{1}{\overline{h}_n^2\,\phi_0^n} \sum_{m=-k_n'}^{k_n'} \overline{\phi}_m^n\, U(|m|)_t^n. \tag{3.8}$$

Note that $\widehat{C}_t^n$ only depends on the variables Y_i^n occurring within the time interval $[0,t]$. The first term on the right hand side above equals C_t, plus a term of order $\Delta_n^{1/4}$ which is asymptotically mixed normal (giving rise to the CLT below), plus a bias term equal to $\frac{\overline{\phi}(0)}{\theta^2} R \int_0^t \gamma_s\, ds$ (with R as in (2.8)); on the other hand, the last term equals this bias term, plus $o(\Delta_n^{1/4})$. Note that for a white noise this bias is the same (with $R = 1$), whether the sampling is regular or not, and in all cases the bias is of order 1: so the de-biasing in (3.8) is necessary, otherwise the estimators would not even be consistent for C_t.

These estimators are consistent for estimating C_t, and enjoy a CLT for any fixed time t. However, to make the CLT feasible, we need consistent estimators for the (conditional) variance in this CLT. Toward this aim, we define a series of processes, where below $m, m' \in \{0,\ldots,k_n\}$:

$$\begin{aligned}
\overline{U}(m,m')_t^n &= \sum_{i=0}^{N_t^n-11k_n} (Y_i^n - \overline{Y}_{i+2k_n}^n)(Y_{i+m}^n - \overline{Y}_{i+4k_n}^n)(Y_{i+6k_n}^n - \overline{Y}_{i+8k_n}^n)(Y_{i+m'+6k_n}^n - \overline{Y}_{i+10k_n}^n)\\
V(m)_t^n &= \sum_{i=0}^{N_t^n-\overline{h}_n'-6k_n} (\widetilde{Y}(\overline{h}_n')_i^n)^2\, 1_{\{|\widetilde{Y}(\overline{h}_n')_i^n|\le u_n'\}} \big(Y_{i+h_n'+k_n}^n - \overline{Y}_{i+h_n'+3k_n}^n\big)\big(Y_{i+m+h_n'+k_n}^n - \overline{Y}_{i+h_n'+5k_n}^n\big)\\
V_t^{n,1} &= \sum_{i=0}^{N_t^n-\overline{h}_n'} (\widetilde{Y}(\overline{h}_n')_i^n)^4\, 1_{\{|\widetilde{Y}(\overline{h}_n')_i^n|\le u_n'\}}\\
V_t^{n,2} &= \sum_{m=-k_n'}^{k_n'} V(|m|)_t^n, \qquad\qquad V_t^{n,3} = \sum_{m,m'=-k_n'}^{k_n'} \overline{U}(|m|,|m'|)_t^n.
\end{aligned}$$

Finally we set

$$\Sigma_t^n = \frac{4}{\phi(0)^2}\Big(\frac{\overline{h}_n}{\overline{h}_n'^2}\,\frac{\Phi_{00}}{3\phi(0)^2}\, V_t^{n,1} + \frac{1}{\overline{h}_n\overline{h}_n'}\,\frac{2\Phi_{01}}{\phi(0)}\, V_t^{n,2} + \frac{1}{\overline{h}_n^3}\,\Phi_{11} V_t^{n,3}\Big). \tag{3.9}$$

For our main result below, we also briefly recall the notion of stable convergence in law. We have a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a sequence V_n of real-valued variables on it. An *extension* of this space is $(\widetilde{\Omega},\widetilde{\mathcal{F}},\widetilde{\mathbb{P}})$, where $\widetilde{\Omega} = \Omega\times\Omega'$ and $\widetilde{\mathcal{F}} = \mathcal{F}\otimes\mathcal{F}'$ for some other measurable space $(\Omega',\mathcal{F}')$, whereas $\widetilde{\mathbb{P}}$ is a probability measure on $(\widetilde{\Omega},\widetilde{\mathcal{F}})$ such that $\widetilde{\mathbb{P}}(A\times\Omega') = \mathbb{P}(A)$ for any $A\in\mathcal{F}$. Then, if V is a variable on $(\widetilde{\Omega},\widetilde{\mathcal{F}},\widetilde{\mathcal{P}})$ and $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$, we say that V_n converges $\mathcal{G}$-stably in law to V if $\mathbb{E}(f(V_n)Z) \to \widetilde{\mathbb{E}}(f(V)Z)$ for any continuous bounded f and any bounded $\mathcal{G}$-measurable Z. This is stronger than the ordinary convergence in law, and $\mathcal{G}$-stable convergence implies $\mathcal{G}'$-stable convergence for any sub-σ-field $\mathcal{G}'$ of $\mathcal{G}$. The main property of the $\mathcal{G}$-stable convergence of V_n is that, if further another sequence V_n' of variables on $(\Omega,\mathcal{F},\mathbb{P})$ converges in probability to a $\mathcal{G}$-measurable limit V', then the pair (V_n,V_n') converges $\mathcal{G}$-stably in law to (V,V').

Theorem 3.1. *Assume (H-r), (O-ρ,ρ') and (N-v), with $v > 3$, and $r < \frac{[v]-4}{2[v]-3}$ and $\rho > 1/4$ and $\rho' < 1/2$. Let $\overline{h}_n, \overline{h}_n', k_n, k_n'$ be as in (3.6) and u_n, u_n' as in (3.7). Then, for any $t > 0$, the sequence $\frac{1}{\Delta_n^{1/4}}(\widehat{C}_t^n - C_t)$ converges $\mathcal{F}_\infty$-stably in law to a limiting variable defined on an extension of the original space, and which is of the form*

$$Y_t = \int_0^t \beta_s\, dB_s, \tag{3.10}$$

where B is a standard Wiener process independent of $\mathcal{F}$ and β_t is the square-root of

$$\beta_t^2 = \frac{4}{\phi(0)^2}\Big(\Phi_{00}\,\frac{\theta\sigma_t^4}{\alpha_t} + 2\Phi_{01}\,\frac{\sigma_t^2\gamma_t^2}{\theta}\, R + \Phi_{11}\,\frac{\gamma_t^4\alpha_t}{\theta^3}\, R^2\Big) \tag{3.11}$$

(recall (2.8) for R). Moreover,

$$\frac{1}{\sqrt{\Delta_n}}\,\Sigma_t^n \xrightarrow{\mathbb{P}} \int_0^t \beta_s^2\, ds, \tag{3.12}$$

and therefore, for any $t > 0$, the sequence $(\widehat{C}_t^n - C_t)/\sqrt{\Sigma_t^n}$ converges $\mathcal{F}_\infty$-stably in law to an $\mathcal{N}(0,1)$ variable, independent of $\mathcal{F}$.

Remark 3.2. We have stated the $\mathcal{F}_\infty$-stable convergence in law only, which is enough for us. However, it is likely that the usual stable convergence in law holds as well. □

Remark 3.3. In addition to the truncation threshold u_n, the estimator in (3.8) involves three tuning parameters, $\overline{h}_n$, k_n and k'_n. Alternatively, one can use the following two-scale version of PAV that involves only one tuning parameter $\overline{h}_n$:

$$\widehat{C}_t^{n,TS} = -\frac{1}{3}\left(\frac{1}{\overline{h}_n\,\phi_0^n}\sum_{i=0}^{N_t^n-\overline{h}_n}(\widetilde{Y}(\overline{h}_n)_i^n)^2\,1_{\{|\widetilde{Y}(\overline{h}_n)_i^n|\le u_n\}} - \frac{4}{2\overline{h}_n\,\phi_0^n}\sum_{i=0}^{N_t^n-2\overline{h}_n}(\widetilde{Y}(2\overline{h}_n)_i^n)^2\,1_{\{|\widetilde{Y}(2\overline{h}_n)_i^n|\le u_n\}}\right). \tag{3.13}$$

Compared with $\widehat{C}_t^n$, the alternative estimator $\widehat{C}_t^{n,TS}$ has one fewer tuning parameter hence may be more convenient to use in practice. However, it has a larger asymptotic variance as can be clearly seen via simulations, which is why we do not pursue this estimator in this article. The reason behind the larger asymptotic variance is basically that the bias in $\frac{1}{\overline{h}_n\phi_0^n}\sum_{i=0}^{N_t^n-\overline{h}_n}(\widetilde{Y}(\overline{h}_n)_i^n)^2$ can be corrected by $\frac{1}{\overline{h}_n^2\phi_0^n}\sum_{m=-k'_n}^{k'_n}\overline{\phi}_m^n\,U(|m|)_t^n$, which, (essentially) by results in Jacod et al. (2017), has a higher convergence rate of $\sqrt{n}$, and consequently does not cause efficiency loss.

Remark 3.4. If we do not truncate in (3.8), then we get consistent estimates of the quadratic variation of X even when X jumps (and a CLT would still be available in this case, although we do not pursue it in this paper).

Remark 3.5. Because white noise is a special case of our setting, Theorem 3.1 implies that when the noise is white, the traditional PAV is still consistent even when observation times are irregularly spaced.

4. Simulation studies

Throughout this section we focus on the time interval [0, 1], which represents one day.

4.1. *Examining Theorem 3.1*

We simulate the case when the underlying process X follows a Heston process (Heston, 1993) with jumps:

$$\begin{cases} dX_t = (\mu - \nu_t/2)dt + \sigma_t\,dW_t + dJ_t, \quad t\in[0,1],\\ d\nu_t = \kappa(\alpha-\nu_t)dt + \zeta\,\nu_t^{1/2}\,dB_t, \end{cases} \tag{4.14}$$

where $\nu_t = \sigma_t^2$, (W_t) and (B_t) are standard Brownian motions with $\mathrm{Corr}(B,W) = \rho$; J is a compound Poisson process independent of (W_t) and (B_t) as follows

$$J_t = \sum_{i=0}^{N_t} D_i,$$

where (N_t) is a Poisson process with rate λ, and D_i's are i.i.d. symmetric mixed normals: $D_i = B_i\cdot Z'_i$, where B_i takes values 1 and -1 with equal probability 0.5, and $Z'_i \sim N(\mu', \sigma'^2)$.

As to the noise process, in the empirical studies in Jacod et al. (2017), we found that the autocorrelations of the microstructure noise decay slowly. For this reason, we choose χ to be a moving-average (MA) series which approximates a fractionally differenced process. More specifically, for an exponent $d\in(-0.5, 0.5)$ and for a large cutoff value M, χ admits the following representation

$$\chi_i = Z''_i + \sum_{j=1}^{M}\psi_j Z''_{i-j}, \quad \text{where } \psi_j = \frac{d(1+d)\cdots(j-1+d)}{j!}, \text{ and } Z''_i \sim_{\text{i.i.d.}} N(0,\sigma_0^2). \tag{4.15}$$

If one lets $M=\infty$, then one obtains a fractionally differenced process whose autocorrelation function decays polynomially at the rate of $2d-1$; see pp. 72–73 in Tsay (2002). For finite but large M, the autocorrelation function first decays slowly till lag M after which the autocorrelation vanishes. We have an additional process γ, which we assume to be an Ornstein–Uhlenbeck-type process featuring a U-shaped pattern as follows

$$d\gamma_t = -\rho_\gamma(\gamma_t-\mu_t)\,dt + \sigma_\gamma\,dW_t, \tag{4.16}$$

where μ_t is a deterministic process which is U-shaped, and W is the *same* Brownian motion as in (4.14). The motivation for specifying γ to be U-shaped comes from the empirical studies in Jacod et al. (2017) where a (vague) U-shaped pattern is seen when examining the size of the noise within each day. The observation times $T(n,i)$ follow an *inhomogeneous* Poisson process with rate $(1+\cos(2\pi t)/2)\times n$, and the noise is $\varepsilon_i^n = \gamma_{T(n,i)}\chi_i$. Note that in this case the noise ε_i^n is *dependent* on the underlying process X. The observations are $Y_{T(n,i)} = X_{T(n,i)} + \varepsilon_i^n,\ i=0,\ldots.$

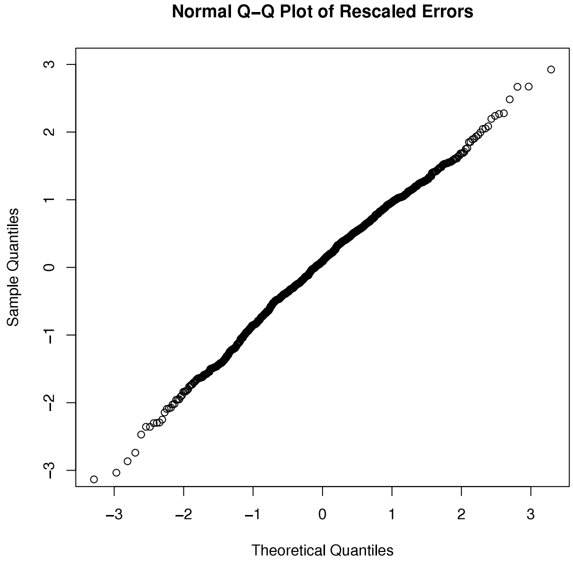

Fig. 1. QQ-plot of normalized errors $(\hat{C}^n_1 - C_1)/\sqrt{\Sigma^n_1}$.

The parameters (4.14), (4.15) and (4.16) are taken as follows:

$$\begin{aligned}
&\mu = 0.05/252, \quad \kappa = 5/252, \quad \alpha = 0.04/252, \quad \zeta = 0.05/252, \quad \rho = -0.5\\
&\lambda = 3, \quad \mu' = 0.001, \quad \sigma' = 0.0003,\\
&d = 0.3, \quad M = 100, \quad \sigma_0 = 0.0005,\\
&\rho_\gamma = 10, \quad \mu_t = 1 + 0.1\cos(2\pi t), \quad \text{and} \quad \sigma_\gamma = 0.1.
\end{aligned} \tag{4.17}$$

The parameter n for the inhomogeneous Poisson process is 23,400. When estimating the integrated volatility, the tuning parameters θ, h_n, h'_n, k_n and k'_n in (3.6) are set to be

$$\overline{h}_n = [0.5\sqrt{N}],\ \overline{h}'_n = [0.8N^{0.7}],\ k_n = [N^{1/5}],\ \text{and}\ k'_n = [N^{1/8}], \tag{4.18}$$

where $N = N^n_1$ as in (2.3).

We check in Fig. 1 the QQ-plot of normalized errors, i.e., $(\hat{C}^n_1 - C_1)/\sqrt{\Sigma^n_1}$, for $\hat{C}^n_t$ and Σ^n_t given respectively by (3.8) and (3.9), based on 1000 replications. The standard normality is clearly supported. A Shapiro–Wilk test also does not reject normality.

4.2. *Comparisons with the traditional pre-averaging estimator*

We compare the performance of our estimator $\widehat{C}^n_1$ with the pre-averaging (PAV) method of Jacod et al. (2009). We pick PAV as the benchmark because it is rate-optimal and has been shown in various applications to perform well.

In the following, since PAV was developed under the equidistant time assumption and when the price process is continuous (although it also works when there are jumps, upon a proper truncation), we shall adopt an equidistant time setting and omit the jump term in (4.14) for comparisons. More specifically, throughout this subsection, the observation times are $T(n,i) = i/n$ for $n = 23{,}400$.

4.2.1. *When Assumption (NO) is satisfied*

We consider the same setting for X, γ and χ as above, the same parameter values as in (4.17), and also the same tuning parameters $\overline{h}_n$, k_n and k'_n as in (4.18) ($\overline{h}'_n$ will not be needed as it is only used in estimating the standard error). Namely, $(\overline{h}_n, k_n, k'_n) = (76, 7, 4)$ in this setting.

As to PAV, it only involves one tuning parameter, that is the width of moving average $\theta\sqrt{n}$. In the study below θ is set to be 0.5, just as in Aït-Sahalia et al. (2013). We have also tried $\theta = 1/3$ as in the numerical studies of Jacod et al. (2009) and the "optimal" θ in the sense of Remark 2 of Jacod et al. (2009). These two choices resulted in larger bias and RMSE and are omitted in the table. Notice that the "optimal" θ may not be optimal because the volatility is stochastic, as discussed in Remark 2 of Jacod et al. (2009).

Table 1 documents the results.

We see that our estimator $\widehat{C}^n_1$ yields both smaller bias and smaller RMSE. The bias of PAV is about 30% of the target (the underlying integrated volatility is about 1.6e−04) and is of the same order as the corresponding RMSE. In contrast, $\widehat{C}^n_1$ has a substantially smaller bias. The RMSE of $\widehat{C}^n_1$ is about 80% smaller than that of the PAV.

Table 1
Comparisons of our estimator $\widehat{C}_1^n$ with PAV based on 1000 replications.

	PAV	$\widehat{C}_1^n$
Bias	5.31e−05	1.75e−06
RMSE	5.44e−05	1.09e−05

Table 2
Comparisons of our estimator $\widehat{C}_1^n$ with PAV based on 1000 replications, when there is rounding.

	PAV	$\widehat{C}_1^n$
Bias	5.31e−05	1.77e−06
RMSE	5.45e−05	1.10e−05

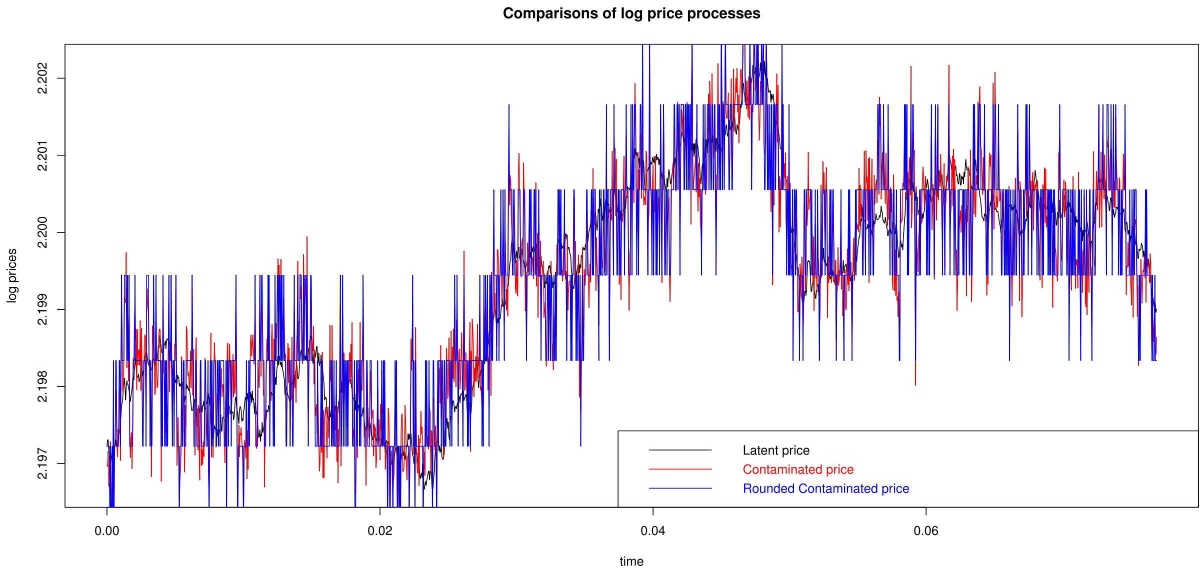

Fig. 2. Comparison of three (log-)price processes, including the latent price process, the contaminated price process, and the rounded prices as in (4.19).

4.2.2. *When there is rounding*

We further investigate a case which is closer to reality but with the assumptions of our estimator violated. More specifically, we study what happens when rounding is involved. We consider the same setting as before, except that the observed prices are the (contaminated) observed prices in the previous subsection further rounded to cents, or equivalently, the observed log prices are

$$Y_i^n = \log([e^{(X_{T(n,i)}+\varepsilon_i^n)}/0.01] * 0.01). \tag{4.19}$$

A sample price process is plotted in Fig. 2.

One can clearly see the rounding effect from Fig. 2. Note that the "noise" here, which is $Y_i^n - X_{T(n,i)}$, does *not* satisfy Assumption (N-v).

Table 2 gives the comparison results.

We see similar comparisons as before, namely, our estimator $\widehat{C}_1^n$ yields substantially smaller bias and RMSE.

We also see in the following QQ plot that the standard normality of the normalized errors still roughly holds (see Fig. 3).

5. Empirical studies

In this section, we apply our "Modified PAV" $\widehat{C}_1^n$ and the traditional PAV to estimate the daily volatilities of the stock Citigroup Inc (NYSE:C), over the month of January 2011. The average size of complete tick-by-tick data is about 246,000 per day. Following Aït-Sahalia et al. (2011), we remove the outliers due to bouncebacks at the cutoff level of 1%; see Section 7.5 therein. After removing such outliers, there are no more apparent jumps. Because the traditional PAV does not involve truncation, for comparison purpose, we make no truncation in implementing both the traditional PAV and our estimator $\widehat{C}_1^n$. Moreover, as we pointed out in Remark 3.5, when the noise is white, the traditional PAV is still consistent even when

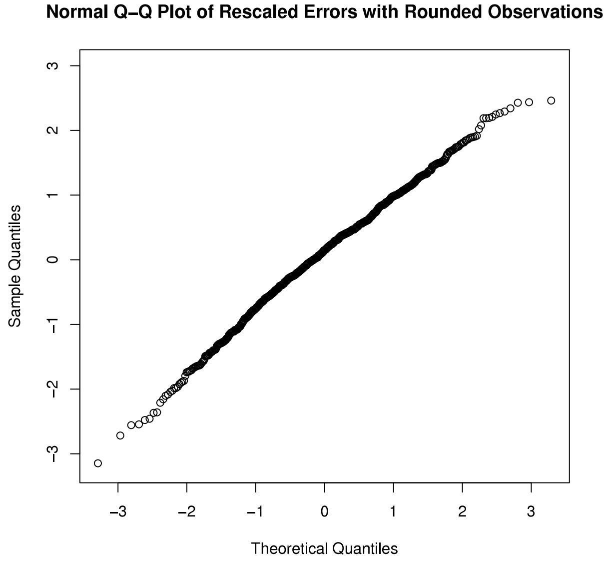

Fig. 3. QQ-plot of normalized errors $(\hat{C}_1^n - C_1)/\sqrt{\Sigma_1^n}$ when observations involve rounding.

observation times are irregularly spaced. For this reason, when applying PAV to tick-by-tick data, we just ignore the fact that the times are irregularly spaced and apply the estimator as usual.

There are a total of 20 trading days in January 2011. For each day, we apply the two estimators to (1) the complete tick-by-tick data and (2) data sampled every one second.

About the choices of the tuning parameters, same as in the simulation studies above, the θ for PAV is set to be 0.5 (we also tried the "optimal" θ in the sense of Remark 2 of Jacod et al. (2009), and the results turn out to be worse). For our estimator, it involves three tuning parameters, namely, $\overline{h}_n$, k_n and k'_n in (3.6) ($\overline{h}'_n$ is only used to estimate the standard error). In the study below, $\overline{h}_n$ is chosen to be the same as the corresponding value for PAV, namely $0.5\sqrt{N}$. As to (k_n, k'_n), for complete data, $(k_n, k'_n) = (40, 10)$. For 1-second data, $(k_n, k'_n) = (7, 4)$, the same as in the simulation studies when $n = 23{,}400$.

The estimation results are given in Fig. 4.

Several comments about Fig. 4 are in order.

(i) The traditional PAV, when applied to the tick-by-tick data can give rather different estimates than when applied to the 1-second data. This is certainly an undesirable feature and raises a difficult question of which frequency is more appropriate to use in order to estimate the latent daily volatility.

(ii) In contrast, the modified PAV gives much more stable estimates.

(iii) The traditional PAV yields higher estimated volatility than the modified PAV. We believe the reason behind such a difference is that the microstructure noise for this particular stock is positively autocorrelated (see Jacod et al. (2017)), and such a positive autocorrelation induces a positive bias in the traditional PAV estimator.

(iv) If we consider 95% confidence intervals, for the traditional PAV, its estimates based on tick-by-tick data are outside the 95% confidence intervals built using the 1-second data for 14 (=70%) out of 20 trading days under study. This proportion is suspiciously high. For our modified PAV, the proportion is 25%. Note that because the estimates based on tick-by-tick data still involve estimation errors, the non-coverage is expected to be higher than 5%.

On the other hand, if we compare the aforementioned estimators with the realized volatility (RV) based on 5-min data, we find that RV yields estimates of similar level to the traditional PAV; see the bottom panel of Fig. 4. Note that RV is always *positively biased*. This confirms the positive bias in the traditional PAV when (blindly) applied to tick-to-tick data. The RV based on 1-min data yields significantly higher estimates.

6. Conclusions

We propose an estimator of the integrated volatility under a setting when the observation times can be irregular and endogenous; the market microstructure noise can have polynomially decaying autocorrelations, can be dependent on the price process, can have diurnal features; and the latent price process can have jumps. We establish consistency as well as CLTs for the proposed estimator. Simulation studies show that the estimator performs well, not only when our assumptions are satisfied, but also even when rounding is involved (in which case our assumptions are violated). Empirical studies further demonstrate the advantages of our proposed estimator, mainly, (1) it gives more stable estimates when applied to different sampling frequencies, including the complete tick-by-tick data; and (2) it can correct the (positive) bias due to the (positive) autocorrelations in the market microstructure noise.

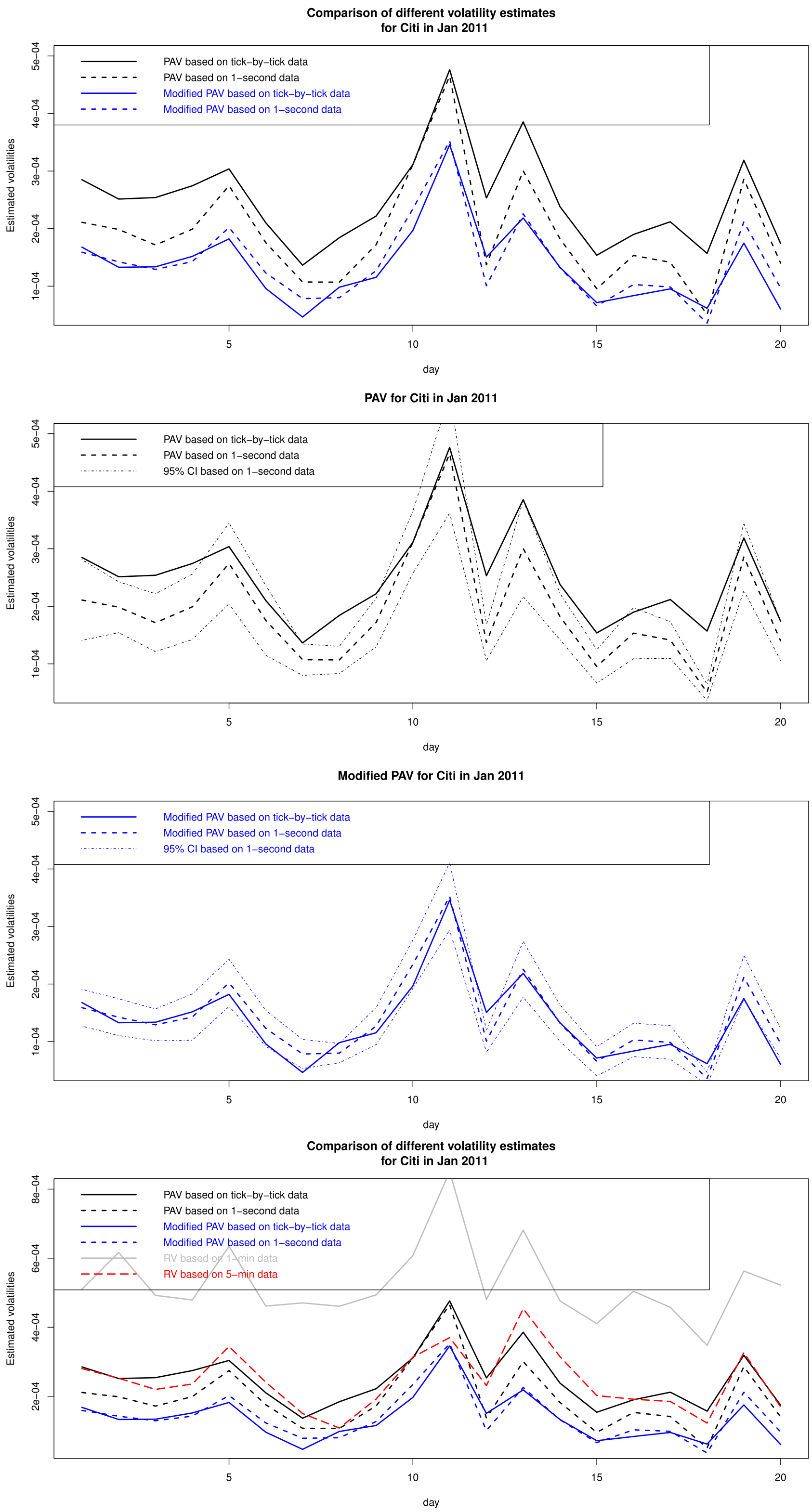

Fig. 4. Comparison of different volatility estimates and the associated 95% confidence intervals for Citigroup Inc. in January 2011.

Acknowledgments

We are very grateful to the editor and two anonymous referees for their valuable and constructive comments.
Yingying Li's research is supported in part by RGC grants GRF 602710, GRF 16502014 and GRF 16518716 of the HKSAR; Xinghua Zheng's research is supported in part by RGC grants GRF 606811, GRF 16305315 and GRF 16304317 of the HKSAR.

Appendix A. Scheme of the proofs

In this appendix we give the proof of the main theorem; it contains a long series of lemmas, which are not proven below, but in the Supplement.

A.1. Preliminaries

Assumption (O-ρ, ρ') becomes weaker when ρ decreases and ρ' increases, hence, in Theorem 3.1 we may assume

$$\frac{1}{4} < \rho \leq \frac{1}{2}, \qquad \frac{1}{4} < \rho' < \frac{1}{2}, \qquad 3 < v \leq \frac{7}{2}. \tag{A.1}$$

Below, we fix a real $\kappa > 2$ arbitrarily large, according to our future needs. Then, upon using a classical localization procedure, there is no restriction to replace the three assumptions (H-r), (O-ρ, ρ') and (N-v) by the following stronger one (in (A.2), the last inequality holds of course for all $\kappa' \in [0, \kappa]$, but unfortunately it seems impossible to "localize" the assumption in such a way that this inequality holds simultaneously for all $\kappa > 0$, even if we allow $K = K_\kappa$ to depend on κ.)

Assumption (S-HON): We have (H-r), (O-ρ, ρ') and (N-v) with $\tau_1 = \infty$ in (O-ρ, ρ') and in Assumption (K-2) for the four processes $b,\ \sigma,\ \alpha,\ \gamma$; moreover the function δ and the processes $b,\ \sigma,\ \alpha,\ 1/\alpha,\ \gamma,\ X$ are bounded and, for some constant $K \geq 1$

$$\begin{array}{l} N_t^n \leq \frac{Kt}{\Delta_n}, \qquad \Big|\mathbb{E}\Big(\Delta(n,i) - \frac{\Delta_n}{\alpha_{T(n,i-1)}} \mid \mathcal{F}_{T(n,i-1)}\Big)\Big| \leq K\Delta_n^{1+\rho} \\ \mathbb{E}\big(\Delta(n,i)^\kappa \mid \mathcal{F}_{T(n,i-1)}\big) \leq K\Delta_n^\kappa \end{array} \tag{A.2}$$

We consider the σ-fields $\mathcal{G}_i = \sigma(\chi_k : j \leq i)$ and $\mathcal{G}^i = \sigma(\chi_k : j \geq i)$ of the "past" and the "future" of the pure noise χ, and the corresponding global σ-field $\mathcal{G} = \mathcal{G}_\infty = \mathcal{G}^{-\infty}$. Assumption (N) implies that $\mathcal{G}$ and $\mathcal{F}_\infty$ are independent. Furthermore, according to (A.4) of Jacod et al. (2017), the ρ-mixing property implies the existence of a constant K such that, for all $j > 0$ and all square-integrable variables ξ and ξ', measurable with respect to $\mathcal{G}^{i+j}$ and $\mathcal{G}_i$ respectively, we have

$$\big|\mathbb{E}(\xi\,\xi' \mid \mathcal{F}_\infty)\big| = \big|\mathbb{E}(\xi\,\xi')\big| \leq \big|\mathbb{E}(\xi)\,\mathbb{E}(\xi')\big| + \frac{K}{j^v}\sqrt{\mathbb{E}(\xi^2)\,\mathbb{E}(\xi'^2)}, \tag{A.3}$$

which, in particular, implies (2.8). Moreover, when ξ is square-integrable centered $\mathcal{G}^{i+j}$-measurable, we have

$$\mathbb{E}\big(\mathbb{E}(\xi \mid \mathcal{F}_\infty \otimes \mathcal{G}_i)^2 \mid \mathcal{F}_\infty\big) \;=\; \mathbb{E}\big(\mathbb{E}(\xi \mid \mathcal{G}_i)^2\big) \;\leq\; \frac{K}{j^{2v}}\,\mathbb{E}(\xi^2). \tag{A.4}$$

We always assume (S-HON) below. Following usual conventions, the constant K appearing below may vary from line to line, and may depend on the bounds occurring in (S-HON) and the kernel function g, but *not* on ω, n or the various indices $i, j, \ldots$ or multi-indices $\mathbf{j}$ (to be defined below); when it depends on additional parameters such as q or t we write it as K_q or $K_{q,t}$. We also use random variables Ψ^n_{par}, usually depending on n and other indices or parameters "*par*" and varying from line to line, but which *always* satisfy

$$\Psi^n_{par} \text{ is nonnegative } \mathcal{G}\text{-measurable with } \mathbb{E}((\Psi^n_{par})^2) \leq 1. \tag{A.5}$$

Recall that, if V is one of the processes $X,\ b,\ \sigma,\ \sigma^2,\ \alpha,\ 1/\alpha,\ \gamma$, or any power of them, for any two finite stopping times $S \leq S'$ and $w \geq 2$, we have

$$\begin{array}{l} \big|\mathbb{E}(V_{S'} - V_S \mid \mathcal{F}_S)\big| \leq K\,\mathbb{E}(S' - S \mid \mathcal{F}_S) \\ \mathbb{E}\Big(\sup_{s\in[S,S']}|V_s - V_S|^w \mid \mathcal{F}_S\Big) \leq K_w\big(\mathbb{E}((S'-S)^w \mid \mathcal{F}_S) + \mathbb{E}(S'-S \mid \mathcal{F}_S)\big). \end{array}$$

Combining this with (A.2) we also get for any $i, j \geq 0$, $s \geq 0$ and $w \in [2, \kappa]$:

$$\begin{array}{l} \big|\mathbb{E}(V_{(T(n,i)+s)\wedge T(n,i+j)} - V_{T(n,i)} \mid \mathcal{F}_{T(n,i)})\big| \leq Kj\Delta_n \\ \mathbb{E}\Big(\sup_{s\in[T(n,i),T(n,i+j)]}|V_s - V_{T(n,i)}|^w \mid \mathcal{F}_{T(n,i)}\Big) \leq K_w\big((j\Delta_n)^w + j\Delta_n\big). \end{array} \tag{A.6}$$

Finally, we need to single out the continuous and purely discontinuous parts of X: we have $X = X^c + X^d$, where

$$X_t^c = X_0 + \int_0^t b_s\,ds + \int_0^t \sigma_s\,dW_s, \qquad X_t^d = \int_{[0,t]\times E} \delta(s,z)\underline{p}(ds,dz). \tag{A.7}$$

The ith observation at stage n is $Y_i^n = Y_i^{c,n} + X^d_{T(n,i)}$, where $Y_i^{c,n} = X^c_{T(n,i)} + \varepsilon_i^n$. Hence the pre-averaged value $\widetilde{Y}(h_n)_i^n$ of (3.3) splits as

$$\widetilde{Y}(h_n)_i^n = \widetilde{Y}(h_n)_i^{c,n} + \widetilde{X}(h_n)_i^{d,n}, \text{ where } \widetilde{Y}(h_n)_i^{c,n} = \sum_{j=1}^{h_n-1} g_j^n\,(Y_{i+j}^{c,n} - Y_{i+j-1}^{c,n}) = -\sum_{j=0}^{h_n-1} \overline{g}_j^n\, Y_{i+j}^{c,n}.$$

A.2. A key decomposition

We need to a study the two pre-averaged values $\widetilde{Y}(\overline{h}_n)_i^n$ and $\widetilde{Y}(\overline{h}'_n)_i^n$. To avoid repeating twice the same arguments, we use a generic sequence h_n, which stands for either $\overline{h}_n$ or $\overline{h}'_n$, and satisfies

$$h_n^{3/2}\Delta_n \to 0, \qquad \inf_n h_n^2\Delta_n > 0. \tag{A.8}$$

The corresponding pre-averaged values for any array V_i^n are simply denoted as $\widetilde{V}_i^n = \widetilde{V}(h_n)_i^n$.

The proof will be similar in spirit to the one in Jacod et al. (2009). We split the sums in the definition of $\widehat{C}_t^n$ into "big" blocks of size ph_n, with p an integer eventually going to ∞, separated by "small" blocks of size $2h_n$, which are eventually negligible but ensure enough conditional independence between the big blocks for having a central limit theorem. We start with a big block and end with a small block. The number of pairs of blocks and the largest index $i = I_n(p,t)$ for the variables $\widetilde{Y}(\overline{h}_n)_i^n$ used inside the last small block are, respectively,

$$J_n(p,t) = 1 + \left[\frac{N_t^n}{(p+2)h_n}\right], \qquad I_n(p,t) = J_n(p,t)(p+2)h_n - 1. \tag{A.9}$$

Notice that the time spanned by the blocks we use is larger than $[0,t]$, but "not too much", and we will need to subtract the superfluous summands, which will turn out to be eventually negligible. The reason for using extra summands is to ensure that $J_n(p,t)$ is a stopping time for a suitable discrete-time filtration.

We begin by defining suitable moments of the "pure" noise χ_i, and for this we need a few specific notation. Let $\mathcal{J}$ be the set of all finite sequences of integers $\mathbf{j} = (j_1, j_2, \ldots, j_q)$ (they are neither necessarily ordered, nor necessarily distinct and $q \geq 1$), and associate the numbers:

$$\mathbf{j} = (j_1, j_2, \ldots, j_q) \ \mapsto \ \ q(\mathbf{j}) = q, \ \ \mu(\mathbf{j}) = \max(j_1, \ldots, j_q).$$

We also denote by $\mathcal{J}^+$ the subset of all $\mathbf{j} \in \mathcal{J}$ whose all entries j_l are nonnegative. We then define for $\mathbf{j} \in \mathcal{J}$ and $m \geq 0$:

$$\begin{aligned} &\mathbf{r}(\mathbf{j}) = \mathbb{E}\Big(\prod_{k=1}^q \chi_{j_k}\Big), \qquad \overline{r}(m)_n = \mathbb{E}\big((\chi_0 - \overline{X}^n_{i+2k_n})(\chi_m - \overline{X}^n_{i+4k_n})\big) \\ &\widetilde{\mathbf{r}}(\mathbf{j})_n = \mathbb{E}\Big(\prod_{k=1}^q \widetilde{\chi}^n_{j_k}\Big) = (-1)^q \sum_{l_1,\ldots,l_q=0}^{h_n-1} \mathbf{r}(j_1 + l_1, \ldots, j_q + l_q) \prod_{k=1}^q \overline{g}^n_{l_k}. \end{aligned} \tag{A.10}$$

We have $\mathbf{r}(\mathbf{j}) = 0$ when $q(\mathbf{j}) = 1$, and $\mathbf{r}(j_1, \ldots, j_q) = \mathbf{r}(m + j_1, \ldots, m + j_q)$ for all $m \in \mathbb{Z}$. Similar properties hold for $\widetilde{\mathbf{r}}(\mathbf{j})$. We have $\mathbf{r}(j) = \widetilde{\mathbf{r}}(j)_n = 0$, and the autocovariance is $r(j) = \mathbf{r}(0, j)$.

Next, we introduce a large number of notation, in connection with the block splitting. Recall $V_i^n = V_{T(n,i)}$ for any process V not depending on n. Below, $p \geq 2$ and $i, j \geq 0$ are integers:

$$\begin{aligned} &\mathcal{F}_i^n = \mathcal{F}_{T(n,i)}, \qquad \mathcal{H}_i^n = \mathcal{F}_\infty \otimes \mathcal{G}_{i-h_n}, \qquad \mathcal{K}_i^n = \mathcal{F}_i^n \otimes \mathcal{G}_{i-h_n} \\ &\qquad \mathcal{H}(p)_j^n = \mathcal{K}^n_{j(p+2)h_n}, \qquad \mathcal{H}'(p)_j^n = \mathcal{K}^n_{(j(p+2)+p)h_n}. \end{aligned}$$

$$\begin{aligned} &\widehat{c}_i^n = \sum_{j\in\mathbb{Z}} (g_j^n)^2 \Delta_{i+j}^n C, \qquad \widehat{X}_i^{c,n} = (\widetilde{X}_i^{c,n})^2 - \widehat{c}_i^n, \\ &\widehat{\varepsilon}_i^n = (\widetilde{\varepsilon}_i^n)^2 - (\gamma_i^n)^2 \widetilde{\mathbf{r}}(0,0)_n, \quad \widehat{X^c\varepsilon}_i^n = \widetilde{X}_i^{c,n}\,\widetilde{\varepsilon}_i^n \end{aligned}$$

$$Z_i^n = (\widetilde{Y}_i^{c,n})^2 - \widehat{c}_i^n - (\gamma_i^n)^2 \widetilde{\mathbf{r}}(0,0)_n = \widehat{X}_i^{c,n} + \widehat{\varepsilon}_i^n + 2\widehat{X^c\varepsilon}_i^n, \quad \zeta(p)_i^n = \sum_{j=i}^{i+ph_n-1} Z_j^n \tag{A.11}$$

$$\begin{aligned} \eta(p)_j^n &= \frac{1}{h_n\phi_0^n}\,\zeta(p)^n_{(j-1)(p+2)h_n}, \qquad \overline{\eta}(p)_j^n = \mathbb{E}(\eta(p)_j^n \mid \mathcal{H}(p)_{j-1}^n), \\ \eta'(p)_j^n &= \frac{1}{h_n\phi_0^n}\,\zeta(p)^n_{(j-1)(p+2)h_n+ph_n}, \quad \overline{\eta}'(p)_j^n = \mathbb{E}(\eta'(p)_j^n \mid \mathcal{H}'(p)_{j-1}^n). \end{aligned}$$

Finally, we define the following processes:

$$\begin{aligned}
F(p)^n_t &= \sum_{j=1}^{J_n(p,t)} \overline{\eta}(p)^n_j, \qquad M(p)^n_t = \sum_{j=1}^{J_n(p,t)} (\eta(p)^n_j - \overline{\eta}(p)^n_j),\\
F'(p)^n_t &= \sum_{j=1}^{J_n(p,t)} \overline{\eta}'(p)^n_j, \qquad M'(p)^n_t = \sum_{j=1}^{J_n(p,t)} (\eta'(p)^n_j - \overline{\eta}'(p)^n_j)\\
\widehat{C}(p)^n_t &= \tfrac{1}{h_n\phi^n_0} \sum_{i=N^n_t-h_n+2}^{I_n(p,t)} Z^n_i, \quad \widehat{C}^{n,1}_t = \tfrac{1}{h_n\phi^n_0} \sum_{i=0}^{N^n_t-h_n} \widehat{c}^n_i - C_t\\
\widehat{C}^{n,2}_t &= \tfrac{1}{h_n\phi^n_0} \sum_{i=0}^{N^n_t-h_n} \Big((\widetilde{Y}^n_i)^2\, 1_{\{|\widetilde{Y}^n_i|\le u_n\}} - (\widetilde{Y}^{c,n}_i)^2\Big)\\
\widehat{C}^{n,3}_t &= \tfrac{\widetilde{\mathbf{r}}(0,0)_n}{h_n\phi^n_0} \sum_{i=0}^{N^n_t-h_n} (\gamma^n_i)^2 - \tfrac{1}{h^2_n\phi^n_0} \sum_{m=-k'_n}^{k'_n} \overline{\phi}^n_m\, U(|m|)^n_t.
\end{aligned} \tag{A.12}$$

Then, for all $p \ge 1$, and with the choice $h_n = \overline{h}_n$ for h_n, we have

$$\widehat{C}^n_t - C_t = M(p)^n_t + M'(p)^n_t + F(p)^n_t + F'(p)^n_t - \widehat{C}(p)^n_t + \sum_{j=1}^{3} \widehat{C}^{n,j}_t, \tag{A.13}$$

and the leading term is $M(p)^n_t + M'(p)^n_t$, all others being residual terms. Note that Z^n_i is $\mathcal{K}^n_{i+2h_n}$-measurable, hence $\zeta(p)^n_i$ is $\mathcal{K}^n_{i+(p+2)h_n}$-measurable. Therefore, we have the following key properties:

$$\eta(p)^n_j \ \text{ is } \mathcal{H}(p)^n_j\text{-measurable}, \qquad \eta'(p)^n_j \ \text{ is } \mathcal{H}'(p)^n_j\text{-measurable}. \tag{A.14}$$

Moreover, $J_n(p,t) \le r$ if and only if $N^n_t < r(p+2)h_n$, which is the same as $T(n, r(p+2)h_n) > t$, hence

$$\text{each } J_n(p,t) \text{ is a stopping time for the discrete-time filtration } (\mathcal{H}(p)^n_j)_{j\ge 0}. \tag{A.15}$$

A.3. Estimates for the process X

It is convenient to introduce the auxiliary processes

$$\begin{aligned}
G^{n,i}_s &= \sum_{j=1}^{h_n-1} g(j/h_n)\, 1_{(T(n,i+j-1),T(n,i+j)]}(s)\\
B^{n,i}_t &= \int_0^t b_u G^{n,i}_u\,du, \quad M^{n,i}_t = \int_0^t \sigma_u G^{n,i}_u\,dW_u \quad X^{d,n,i}_t = \int_{[0,t]\times E} \delta(u,z)\, G^{n,i}_u\, \underline{p}(du,dz)\\
X^{c,n,i}_t &= B^{n,i}_t + M^{n,i}_t, \qquad C^{n,i}_t = \int_0^t \sigma^2_u\,(G^{n,i}_u)^2\,du
\end{aligned}$$

They all vanish for $t \le T(n,i)$, and $G^{n,i}_t = 0$ and all others are constant (in time) for $t \ge T'(n,i) := T(n, i+h_n-1)$. We have

$$\widetilde{X}^n_i = \widetilde{X}^{c,n}_i + \widetilde{X}^{d,n}_i, \qquad \widetilde{X}^{c,n}_i = X^{c,n,i}_{T'(n,i)}, \qquad \widetilde{X}^{d,n}_i = X^{d,n,i}_{T'(n,i)}, \qquad \widehat{c}^n_i = C^{n,i}_{T'(n,i)},$$

and we define for any $j, j' \ge 0$ the variables

$$\widehat{G}^{n,i}_{j,j'} = \int_0^\infty G^{n,i+j}_u G^{n,i+j'}_u\,du, \qquad \overline{G}^{n,i}_{j,j'} = \int_0^\infty G^{n,i+j}_u G^{n,i+j'}_u\,du \int_0^u G^{n,i+j}_s G^{n,i+j'}_s\,ds. \tag{A.16}$$

Upon using (A.2) and Hölder's inequality, and also Burkholder–Gundy inequality for $M^{n,i}$ below, we have for $w \in [0,\kappa]$ and $w' \in [r,\kappa]$ and $0 \le j < j'$:

$$\begin{aligned}
&\mathbb{E}\Big((\mathrm{Var}(B^{n,i}_\infty))^w + (C^{n,i}_\infty)^w \mid \mathcal{F}^n_i\Big) \le K_w(h_n\Delta_n)^w\\
&\mathbb{E}\Big(|M^{n,i}_{T(n,i+j')} - M^{n,i}_{T(n,i+j)}|^w \mid \mathcal{F}^n_{i+j}\Big) \le K_w((j'-j)\Delta_n)^{w/2}\\
&\mathbb{E}\Big(\sup_s |M^{n,i}_s|^w \mid \mathcal{F}^n_{i+j}\Big) + \mathbb{E}\Big(|\widetilde{X}^{c,n}_i|^w \mid \mathcal{F}^n_{i+j}\Big) \le K_w(h_n\Delta_n)^{w/2}.
\end{aligned} \tag{A.17}$$

For $X^{d,n,i}$ we have the following estimates (recall $r < 1$):

Lemma A.1. *If* $0 < q < 1/r$ *and* $w \ge r$ *and* $\eta > 0$ *we have*

$$\mathbb{E}(|(h_n\Delta_n)^{-q}\widetilde{X}^{d,n}_i|^w \wedge 1 \mid \mathcal{F}^n_i)\Big| \le K_{q,w,\eta}(h_n\Delta_n)^{1-qr-\eta}. \tag{A.18}$$

Lemma A.2. *For any* $w \in [2,\kappa/2]$ *and* $0 \le j \le j' \le ph_n$ *we have*

$$\Big|\,\mathbb{E}(\widehat{X}^{c,n}_i \mid \mathcal{F}^n_i)\Big| \le Kh_n\Delta_n^{3/2}(1 + h_n^{1/2}\Delta_n^{\rho}), \qquad \mathbb{E}(|\widehat{X}^{c,n}_i|^w \mid \mathcal{F}^n_i) \le K_w(h_n\Delta_n)^w \tag{A.19}$$

$$\Big|\,\mathbb{E}\Big(\widehat{X}^{c,n}_{i+j}\,\widehat{X}^{c,n}_{i+j'} - 4(\sigma^n_i)^4\,\overline{G}^{n,i}_{j,j'} \mid \mathcal{F}^n_i\Big)\Big| \le K_p\,(h_n\Delta_n)^{5/2} \tag{A.20}$$

$$\Big|\,\mathbb{E}\Big(\widehat{c}^n_i - (\sigma^n_i)^2\,\widehat{G}^{n,i}_{0,0} \mid \mathcal{F}^n_i\Big)\Big| \le K(h_n\Delta_n)^2 \tag{A.21}$$

$$\Big|\,\mathbb{E}\Big((\widetilde{X}^{c,n}_i)^4 - 6(\sigma^n_i)^4\,\overline{G}^{n,i}_{0,0} \mid \mathcal{F}^n_i\Big)\Big| \le K(h_n\Delta_n)^{5/2}. \tag{A.22}$$

A.4. Estimates for noise

Lemma A.3. *Let $p \ge 1$. For all $\mathbf{j} = (j_1,\ldots,j_q) \in \mathcal{J}^+$ with $\mu(\mathbf{j}) \le ph_n$ we have*

$$|\widetilde{\mathbf{r}}(\mathbf{j})_n| \le K_{p,q}/h_n^{v\bigwedge[(q+1)/2]}. \tag{A.23}$$

In connection with the notation (A.5), we now explain a simple procedure which will often be used. Consider two integrable variables, U which is $\mathcal{F}_\infty$-measurable, and ξ which is $\mathcal{G}$-measurable. Then

$$\big|\,\mathbb{E}(\xi \mid \mathcal{G}_{i-h_n})\big| \le K\Psi^n_i \;\;\Rightarrow\;\; \big|\,\mathbb{E}(U\,\xi \mid \mathcal{K}^n_i)\big| \le K\Psi^n_i\,\big|\,\mathbb{E}(U \mid \mathcal{F}^n_i)\big|. \tag{A.24}$$

This follows from the property $\mathbb{E}(U\,\xi \mid \mathcal{K}^n_i) = \mathbb{E}(\xi \mid \mathcal{G}_{i-h_n})\,\mathbb{E}(U \mid \mathcal{F}^n_i)$, which in turn comes from the measurability properties of U and ξ, the definition of $\mathcal{K}^n_i$, and the independence between $\mathcal{F}_\infty$ and $\mathcal{G}$.

Lemma A.4. *Let $p \ge 1$ and $q \ge 2$. For all $\mathbf{j} = (j_1,\ldots,j_q) \in \mathcal{J}^+$ with $\mu(\mathbf{j}) \le ph_n$, all $j \in \{0,\ldots,ph_n\}$, and all $\mathcal{F}_\infty$-measurable variables U, we have*

$$\begin{aligned}
&\Big|\,\mathbb{E}\Big(\prod_{m=1}^q \widetilde{\varepsilon}^n_{i+j+j_m} - (\gamma^n_i)^q\,\widetilde{\mathbf{r}}(\mathbf{j})_n \mid \mathcal{K}^n_i\Big)\Big| \le K_{p,q}\Big(\Psi^n_{i,j,\mathbf{j}}/h_n^v + \Delta_n/h_n^{[\frac{q-1}{2}]}\Big)\\
&\Big|\,\mathbb{E}\Big(U\Big(\prod_{m=1}^q \widetilde{\varepsilon}^n_{i+j+j_m} - (\gamma^n_i)^q\,\widetilde{\mathbf{r}}(\mathbf{j})_n\Big) \mid \mathcal{K}^n_i\Big)\Big|\\
&\qquad \le K_{p,q}\Big(\Psi^n_{i,j,\mathbf{j}}/h_n^v + \Delta_n^{1/2}/h_n^{[\frac{q-1}{2}]+\frac{1}{2}}\Big)\sqrt{\mathbb{E}(U^2 \mid \mathcal{F}^n_i)}
\end{aligned} \tag{A.25}$$

and also

$$\Big|\,\mathbb{E}\Big(U\,\widetilde{\varepsilon}^n_{i+j} \mid \mathcal{K}^n_i\Big)\Big| \le K_p\,\frac{\Psi^n_{i,j}}{h_n^v}\,\mathbb{E}(|U| \mid \mathcal{F}^n_i), \tag{A.26}$$

$$q \text{ an even integer } \Rightarrow\; \mathbb{E}\Big((\widetilde{\varepsilon}^n_i)^q \mid \mathcal{K}^n_i\Big) \le K\Psi^n_{i,j}/h_n^{v\bigwedge\frac{q}{2}}, \tag{A.27}$$

$$\mathbb{E}\Big((\widetilde{Y}^{c,n}_i)^q \mid \mathcal{F}^n_i\Big) \le \begin{cases} K\Psi^n_{i,j}(h_n\Delta_n)^{q/2} & \text{if } 0 \le q \le 2[v] \\ K(h_n\Delta_n)^{q/2} + K\Psi^n_{i,j}/h_n^v & \text{if } q > 2v \text{ is an even integer}. \end{cases} \tag{A.28}$$

Lemma A.5. *With the notation (A.16), we have for $j, j' \in \{0,\ldots,ph_n\}$ and $q \ge 2$ and even integer:*

$$\big|\,\mathbb{E}(\widehat{\varepsilon}^n_i \mid \mathcal{K}^n_i)\big| \le K\Psi^n_i\Delta_n, \qquad \big|\,\mathbb{E}(\widehat{X^c\varepsilon}^n_i \mid \mathcal{K}^n_i)\big| \le K\Psi^n_i\,\Delta_n^{1/2}/h_n^{v-1/2} \tag{A.29}$$

$$\mathbb{E}((\widehat{\varepsilon}^n_i)^q \mid \mathcal{K}^n_i) \le K\Psi^n_i/h_n^{v\bigwedge q}, \qquad \mathbb{E}(|\widehat{X^c\varepsilon}^n_i|^4 \mid \mathcal{K}^n_i) \le K\Psi^n_i\,\Delta_n^2 \tag{A.30}$$

$$\big|\,\mathbb{E}(\widehat{\varepsilon}^n_{i+j}\,\widehat{\varepsilon}^n_{i+j'} \mid \mathcal{K}^n_i) - (\gamma^n_i)^4(\widetilde{\mathbf{r}}(j,j,j',j')_n - (\widetilde{\mathbf{r}}(0,0)_n)^2)\big| \le K_p\Psi^n_i\,\Delta_n^{5/4} \tag{A.31}$$

$$\big|\,\mathbb{E}(\widehat{X^c\varepsilon}^n_{i+j}\,\widehat{X^c\varepsilon}^n_{i+j'} \mid \mathcal{K}^n_i) - (\sigma^n_i\gamma^n_i)^2\,\widetilde{\mathbf{r}}(j,j')_n\widehat{G}^{n,i}_{j,j'}\big| \le K_p\Psi^n_i\,h_n^{1/2}\Delta_n^{3/2} \tag{A.32}$$

Lemma A.6. *Let $u_n \asymp (h_n\Delta_n)^{\varpi}$. Under (3.7) there are a number $\eta > 0$ and a constant K (depending on ϖ and on the sequence u_n) such that, for $(w,z) \in \{(1,1),(1,2),(2,1)\}$,*

$$\mathbb{E}\Big(\big|\,(\widetilde{Y}^n_i)^{2z}\,1_{\{|\widetilde{Y}^n_i \le u_n\}} - (\widetilde{Y}^{c,n}_i)^{2z}\big|^w\Big) \le K(h_n\Delta_n)^{1+\frac{wz}{2}+\eta}. \tag{A.33}$$

Finally, we give estimates on the processes $U(m)^n$, $\overline{U}(m,m')^n$ and $V(m)^n$.

Lemma A.7. *If $0 \le m, m' \le k_n$ we have for some $\eta > 0$:*

$$\begin{aligned}
&\mathbb{E}\Big(\Big| U(m)^n_t - r(m) \sum_{i=0}^{N^n_t - 5k_n} (\gamma^n_i)^2 \Big|\Big) \le K_t \,\frac{\sqrt{k_n}}{\sqrt{\Delta_n}} \\
&\mathbb{E}\Big(\Big| \overline{U}(m,m')^n_t - r(m)r(m') \sum_{i=0}^{N^n_t - 11k_n} (\gamma^n_i)^4 \Big|\Big) \le K_t \,\frac{\sqrt{k_n}}{\sqrt{\Delta_n}} \\
&\mathbb{E}\Big(\Big| V(m)^n_t - r(m) \sum_{i=0}^{N^n_t - \overline{h}'_n - 6k_n} (\gamma^n_i)^2 (\widetilde{Y}(\overline{h}'_n)^{c,n}_i)^2 \Big|\Big) \le K_t \,\overline{h}'^{1+\eta}_n \,\Delta^\eta_n.
\end{aligned} \tag{A.34}$$

A.5. Auxiliary computations

Below, our main aim is to estimate various moments of the variables $\zeta(p)^n_i$ of (A.11).

Lemma A.8. *In the case $h_n = \overline{h}_n$, we have*

$$\big|\mathbb{E}(\zeta(p)^n_i \mid \mathcal{K}^n_i)\big| \le K_p \Psi^n_{i,p}\, \Delta^{1/2}_n, \qquad \mathbb{E}(|\zeta(p)^n_i|^4 \mid \mathcal{K}^n_i) \le K_p \Psi^n_{i,p}\, \Delta^{v/2-2}_n \tag{A.35}$$

$$\big|\mathbb{E}\big((\zeta(p)^n_i)^2 - 4(\sigma^n_i)^4 \rho(p,1)^n_i - (\gamma^n_i)^4 \rho(p,2)^n - 4(\sigma^n_i \gamma^n_i)^2 \rho(p,3)^n_i \mid \mathcal{K}^n_i\big)\big| \le K_p \Psi^n_{i,p}\, \Delta^{1/4}_n \tag{A.36}$$

where

$$\rho(p,1)^n_i = \sum_{j,j'=0}^{ph_n-1} \overline{G}^{n,i}_{j,j'}, \qquad \rho(p,2)^n = \sum_{j,j'=0}^{ph_n-1} (\widetilde{\mathbf{r}}(j,j,j',j')_n - (\widetilde{\mathbf{r}}(0,0)_n)^2),$$
$$\rho(p,3)^n_i = \sum_{j,j'=0}^{ph_n-1} \widetilde{\mathbf{r}}(j,j')_n \widehat{G}^{n,i}_{j,j'}.$$

We now have to evaluate the quantities $\rho(p,k)^n_i$. Toward this aim, we complement the notation (3.1) by setting

$$\overline{\Phi}_{00} = \int_0^1 s\phi(s)^2\,ds, \quad \overline{\Phi}_{11} = \int_0^1 s\overline{\phi}(s)^2\,ds, \quad \overline{\Phi}_{01} = \int_0^1 s\phi(s)\overline{\phi}(s)\,ds. \tag{A.37}$$

Lemma A.9. *We have*

$$\Big|\mathbb{E}(\widehat{G}^{n,i}_{0,0} \mid \mathcal{F}^n_i) - \frac{h_n \Delta_n \phi(0)}{\alpha^n_i}\Big| \le K\big(h_n \Delta^{1+\rho}_n + h^2_n \Delta^2_n\big). \tag{A.38}$$

$$\Big|\mathbb{E}(\overline{G}^{n,i}_{0,0} \mid \mathcal{F}^n_i) - \frac{h^2_n \Delta^2_n \phi(0)^2}{2(\alpha^n_i)^2}\Big| \le K\big(h^2_n \Delta^{2+\rho}_n + h^3_n \Delta^3_n\big). \tag{A.39}$$

$$\Big|\mathbb{E}(\rho(p,1)^n \mid \mathcal{F}^n_i) - \frac{h^4_n \Delta^2_n}{(\alpha^n_i)^2} (p\Phi_{00} - \overline{\Phi}_{00})\Big| \le K_p h^3_n \Delta^2_n. \tag{A.40}$$

$$\Big|\rho(p,3)^n_i - 2\,\frac{h^2_n \Delta_n}{\alpha^n_i} (p\Phi_{01} - \overline{\Phi}_{01}) R\Big| \le K_p h_n \Delta_n. \tag{A.41}$$

Lemma A.10. *We have*

$$\big|\rho(p,2)^n - 4(p\Phi_{11} - \overline{\Phi}_{11}) R^2\big| \le K_p / h_n. \tag{A.42}$$

We end this subsection with a simple but useful version of "Riemann integration":

Lemma A.11. *If V is a càdlàg process, and $p \ge 2$, and l_n is an arbitrary sequence of positive integers with $\Delta_n l_n \to 0$, we have for all $t > 0$:*

$$\Delta_n \sum_{i=0}^{(N^n_t - l_n)^+} V_{T(n,i)} \xrightarrow{\mathbb{P}} \int_0^t V_s\,dA_s, \qquad h_n \Delta_n \sum_{j=0}^{J_n(p,t)} V_{T(n,(j-1)(p+2)h_n)} \xrightarrow{\mathbb{P}} \frac{1}{p+2} \int_0^t V_s\,dA_s. \tag{A.43}$$

A.6. The negligible terms

In this subsection we prove that the residual terms defined in (A.12) are negligible once multiplied by the rate $1/\Delta_n^{1/4}$, in the case $h_n = \overline{h}_n$ (this fails when $h_n = \overline{h}'_n$). Lemma A.6 with $w = z = 1$, plus $N_t^n \le K_t/\Delta_n$, readily give us

$$h_n = \overline{h}_n \;\;\Rightarrow\;\; \Delta_n^{-1/4}\,\widehat{C}_t^{n,2} \;\stackrel{\mathbb{P}}{\longrightarrow}\; 0. \tag{A.44}$$

Lemma A.12. *As $n \to \infty$, and if $h_n = \overline{h}_n$, for all $p \ge 2$ and t we have $\Delta_n^{-1/4} F(p)_t^n \stackrel{\mathbb{P}}{\longrightarrow} 0$ and $\Delta_n^{-1/4} F'(p)_t^n \stackrel{\mathbb{P}}{\longrightarrow} 0$.*

Lemma A.13. *As $n \to \infty$, and if $h_n = \overline{h}_n$, for all $p \ge 2$ and t we have $\Delta_n^{-1/4}\,\widehat{C}(p)_t^n \stackrel{\mathbb{P}}{\longrightarrow} 0$.*

Lemma A.14. *As $n \to \infty$, and if $h_n = \overline{h}_n$, for all t we have $\Delta_n^{-1/4}\,\widehat{C}_t^{n,1} \stackrel{\mathbb{P}}{\longrightarrow} 0$.*

Lemma A.15. *As $n \to \infty$, for all t we have $\Delta_n^{-1/4}\,\widehat{C}_t^{n,3} \stackrel{\mathbb{P}}{\longrightarrow} 0$.*

By virtue of (A.13), (A.44), and of the previous three lemmas, the CLT in Theorem 3.1 amounts to the convergence of $\frac{1}{\Delta_n^{1/4}}(M(p)_t^n + M'(p)_t)$. As a matter of fact, at the end we will let $p \to \infty$, so the processes $M'(p)^n$ are also eventually negligible. To substantiate this claim, we state a lemma:

Lemma A.16. *When $h_n = \overline{h}_n$ we have $\mathbb{E}\big(|\Delta_n^{-1/4}\,M'(p)_s^n|^2\big) \le Kt/p$.*

A.7. Proof of Theorem 3.1

We begin with the CLT for $M(p)^n$, which goes as follows (note that we have the "functional" convergence here):

Proposition A.17. *For any fixed $p \ge 2$, and if $h_n = \overline{h}_n$, the sequence $\Delta_n^{-1/4} M(p)^n$ of processes converges $\mathcal{F}_\infty$-stably in law to the process*

$$Y(p)_t \;=\; \int_0^t \beta(p)_s\,dB_s, \tag{A.45}$$

where B is as in Theorem 3.1 and $\beta(p)_t$ is the square root of

$$\beta(p)_t^2 = \frac{4}{\phi(0)^2}\Big(\frac{p\,\Phi_{00} - \overline{\Phi}_{00}}{p+2}\,\frac{\theta\sigma_t^4}{\alpha_t} + 2\Big(\frac{p\,\Phi_{01} - \overline{\Phi}_{01}}{p+2}\Big)\frac{\sigma_t^2\gamma_t^2}{\theta}\,R + \frac{p\,\Phi_{11} - \overline{\Phi}_{11}}{p+2}\,\frac{\gamma_t^4\alpha_t}{\theta^3}\,R^2\Big).$$

This is Lemma 5.7 of Jacod et al. (2009), in a much more general setting and somewhat modified notation, and also the difference that we only prove the $\mathcal{F}_\infty$-stable convergence and not the $\mathcal{F}$-stable convergence.

We set $\widehat{\eta}(p)_j^n = \eta(p)_j^n - \overline{\eta}(p)_j^n$, and $\Delta(V,p)_j^n = V_{j(p+2)h_n} - V_{(j-1)(p+2)h_n}$ for any process V. We also set $\mathcal{M} = \mathcal{M}_1 \cup \{W\}$, where W is the Brownian motion driving X and $\mathcal{M}_1$ denotes the class of all bounded $(\mathcal{F}_t)$-martingales orthogonal to W. By (A.14), (A.15) and a standard limit theorem for triangular arrays of martingale differences, for proving Proposition A.17 it suffices to show the following three convergences:

$$t > 0 \;\Longrightarrow\; \frac{1}{\sqrt{\Delta_n}} \sum_{j=1}^{J_n(p,t)} \mathbb{E}\big((\widehat{\eta}(p)_j^n)^2 \mid \mathcal{H}(p)_{j-1}^n\big) \;\stackrel{\mathbb{P}}{\longrightarrow}\; \int_0^t \beta(p)_s^2\,ds \tag{A.46}$$

$$t > 0 \;\Longrightarrow\; \frac{1}{\Delta_n} \sum_{j=1}^{J_n(p,t)} \mathbb{E}\big((\widehat{\eta}(p)_j^n)^4 \mid \mathcal{H}(p)_{j-1}^n\big) \;\stackrel{\mathbb{P}}{\longrightarrow}\; 0 \tag{A.47}$$

$$t > 0,\; V \in \mathcal{M} \;\Longrightarrow\; \frac{1}{\Delta_n^{1/4}} \sum_{j=1}^{J_n(p,t)} \mathbb{E}\big(\widehat{\eta}(p)_j^n\,\Delta(V,p)_j^n \mid \mathcal{H}(p)_{j-1}^n\big) \;\stackrel{\mathbb{P}}{\longrightarrow}\; 0. \tag{A.48}$$

Proof of (A.46). The jth summand in the left side of (A.46) is $\mathbb{E}\big((\eta(p)_j^n)^2 \mid \mathcal{H}(p)_{j-1}^n\big) - (\overline{\eta}(p)_j^n)^2$, and (A.35) yields $(\overline{\eta}(p)_j^n)^2 \le K_p(\Psi_j'^n)^2\Delta_n^2$. Since $J_n(p,t) \le Kt/\sqrt{\Delta_n}$ and $\mathbb{E}((\Psi_j'^n)^2) \le 1$, we deduce

$$\mathbb{E}\Big(\frac{1}{\sqrt{\Delta_n}} \sum_{j=1}^{J_n(p,t)} (\overline{\eta}(p)_j^n)^2\Big) \le K_p\,t\,\Delta_n \;\to\; 0.$$

Therefore, it is enough to show the convergence when $\widehat{\eta}(p)^n_j$ is substituted with $\eta(p)^n_j$. Moreover, (A.36) and Lemmas A.9 and A.10 imply

$$\Big|\mathbb{E}\big((\eta(p)^n_j)^2 \mid \mathcal{H}(p)^n_{j-1}\big) - \tfrac{4}{(\phi^n_0)^2}\Big(h_n^2\Delta_n^2\,\tfrac{\sigma^4_{T(n,(j-1)(p+2)h_n)}}{\alpha^2_{T(n,(j-1)(p+2)h_n)}}\,(p\Phi_{00}-\overline{\Phi}_{00})$$
$$+\tfrac{1}{h_n^2}\,\gamma^4_{T(n,(j-1)(p+2)h_n)}(p\Phi_{11}-\overline{\Phi}_{11}) + 2\Delta_n\,\tfrac{(\sigma^2\gamma^2)_{T(n,(j-1)(p+2)h_n)}}{\alpha_{T(n,(j-1)(p+2)h_n)}}\,(p\Phi_{01}-\overline{\Phi}_{01})\Big)\Big|$$
$$\leq K_p\,\Psi^n_{j,p}\Delta_n^{5/4}.$$

Hence, the result follows from (A.43), $\phi^n_0 \to \phi(0)$ and $h_n\sqrt{\Delta_n} \to \theta$ because $h_n = \overline{h}_n$. □

Proof of (A.47). It suffices to prove the result with $\widehat{\eta}(p)^n_j$ substituted with $\eta(p)^n_j$. (A.35) and $1/|\phi^n_0| \leq K$ yield $\mathbb{E}\big((\eta(p)^n_j)^4 \mid \mathcal{H}(p)^n_{j-1}\big) \leq K_p\Psi^n_{j,p}\Delta_n^{v/2}$. Using again $J_n(p,t) \leq Kt/\sqrt{\Delta_n}$ and $\mathbb{E}(\Psi^n_{j,p}) \leq 1$, plus $v > 3$, we obtain the result. □

Proof of (A.48). (1) Since $\mathbb{E}(\Delta(V,p)^n_j \mid \mathcal{H}(p)^n_{j-1}) = 0$, the desired convergence for $V \in \mathcal{M}$ amounts to

$$\Delta_n^{1/4}\sum_{j=1}^{J_n(p,t)}\mathbb{E}\big(\zeta(p)^n_{(j-1)(p+2)h_n}\Delta(V,p)^n_j \mid \mathcal{H}(p)^n_{j-1}\big) \stackrel{\mathbb{P}}{\longrightarrow} 0.$$

Recall the decomposition $\widehat{X}^{c,n}_i = 2\sum_{l=1}^4 D^{n,i}_l$ of the proof of Lemma A.2. Setting

$$Z(k)^n_j = \sum_{i=(j-1)(p+2)h_n}^{(j-1)(p+2)h_n+ph_n-1} z(k)^n_i, \quad \text{where}$$
$$z(1)^n_i = 2D^{n,i}_1, \quad z(2)^n_i = 2(D^{n,i}_2 + D^{n,i}_3 + D^{n,i}_4)$$
$$z(3)^n_i = \widehat{\varepsilon}^n_i, \qquad z(4)^n_i = \widehat{X^c\varepsilon}^n_i,$$

we are thus left to proving that, for $k = 1,2,3,4,$

$$A(k)_n := \Delta_n^{1/4}\sum_{j=1}^{J_n(p,t)}\mathbb{E}\big(Z(k)^n_j\Delta(V,p)^n_j \mid \mathcal{H}(p)^n_{j-1}\big) \stackrel{\mathbb{P}}{\longrightarrow} 0. \tag{A.49}$$

(2) For the cases $k = 2,3,4$, in view of (S.24) it suffices to prove (A.2) in restriction to the set $\{\overline{M}^n_{t+1} \leq \Delta^a_n\}$, for some $a \in [1/2+\rho', 1)$. Observing that $(1+(p+2)h_n)\Delta^a_n \leq 1$ (for n large enough), on this set we have $T(n, J_n(p,t)(p+2)h_n) \leq t+1$, hence the left side of (A.48) only involves the process V stopped at time $t+1$. In other words, it is enough to prove the result when V is a square-integrable martingale up to infinity (of course when $V \in \mathcal{M}_1$ this is automatically the case, but this allows us to accommodate the case $V = W$ below, which is thus replaced by $V_s = W_{s\wedge(t+1)}$).

Below, we suppose that i is in the range of the sum defining the variables $Z(k)^n_j$, and we heavily use $h_n = \overline{h}_n \asymp 1/\sqrt{\Delta_n}$. When $k = 2$ we use the estimates (S.1) to obtain

$$\big|\mathbb{E}(z(k)^n_i\Delta(V,p)^n_j \mid \mathcal{H}(p)^n_{j-1})\big| \leq K_p\,\Gamma(k)^n_i\,\sqrt{\mathbb{E}\big((\Delta(V,p)^n_j)^2 \mid \mathcal{H}(p)^n_{j-1}\big)}, \tag{A.50}$$

with $\Gamma(2)^n_i = \Delta_n^{3/4}$. When $k = 3$ we use (A.25) with $U = \Delta(V,p)^n_j$ to get (A.50) with now $\Gamma(3)^n_i = \Psi^n_i\Delta_n^{3/4}$ (use also $v > 3$). When $k = 4$ we use (A.26) with $U = \widetilde{X}^{c,n}_i\Delta(V,p)^n_j$, plus (A.17), to obtain (A.50) once more, with $\Gamma(4)^n_i = \Psi^n_i\Delta_n^{1/4+v/2}$.

One deduces from (A.50) and the definition of $Z(k)^n_j$ that

$$\big|\mathbb{E}(Z(k)^n_j\Delta(V,p)^n_j \mid \mathcal{H}(p)^n_{j-1})\big| \leq K_p\,\Gamma'(k)^n_j\,\sqrt{\mathbb{E}\big((\Delta(V,p)^n_j)^2 \mid \mathcal{H}(p)^n_{j-1}\big)},$$
$$\text{where } \Gamma'(k)^n_j = \sum_{i=(j-1)(p+2)h_n}^{(j-1)(p+2)h_n+ph_n-1}\Gamma(k)^n_i,$$

and the Cauchy–Schwarz inequality yields $\mathbb{E}(A(k)_n)^2 \leq K_p\,\beta_n\delta(k)_n$, where

$$\beta_n = \mathbb{E}\Big(\sum_{j=1}^{J_n(p,t)}(\Delta(V,p)^n_j)^2\Big), \qquad \delta(k)_n = \sqrt{\Delta_n}\,\mathbb{E}\Big(\sum_{j=1}^{J_n(p,t)}(\Gamma'(k)^n_j)^2\Big).$$

Since V is a square-integrable martingale, $\beta_n = \mathbb{E}((V_{T(n,J_n(p,t)(p+2)h_n)} - V_0)^2)$, which is smaller than the finite number $\mathbb{E}((V_\infty - V_0)^2)$. On the other hand,

$$\delta(k)_n \leq ph_n\sqrt{\Delta_n}\,\mathbb{E}\Big(\sum_{i=0}^{J_n(p,t)(p+2)h_n}(\Gamma(k)^n_i)^2\Big) \leq K_p\,\mathbb{E}\Big(\sum_{i=0}^{[\Lambda/\Delta_n]}(\Gamma(k)^n_i)^2\Big),$$

where Λ is a constant (depending on p and t) such that $J_n(p,t)(p+2)h_n \leq \Lambda/\Delta_n$. Since $\mathbb{E}((\Psi_i^n)^2) \leq 1$, and in view of the definition of $\Gamma(k)_i^n$, we deduce that $\delta(k)_n$ is smaller than $K_{p,t}\sqrt{\Delta_n}$ in all cases $k = 2, 3, 4$, so (A.49) holds.

(3) Finally, consider the case $k = 1$. Since $M^{n,i}$ is a stochastic integral with respect to W, when $V \in \mathcal{M}_1$ we see that $\mathbb{E}\big(z(1)_i^n \Delta(V,p)_j^n \mid \mathcal{H}(p)_{j-1}^n\big) = 0$ when i is between $(j-1)(p+2)h_n$ and $(j-1)(p+2)h_n + ph_n - 1$, so $A(1)_n = 0$.

Suppose now that $V = W$, and let i be as just above. Then

$$\mathbb{E}\big(z(1)_i^n \Delta(V,p)_j^n \mid \mathcal{F}_i^n\big) = 2\mathbb{E}\Big(\int_{T(n,i)}^{T'(n,i)} X_s^{c,n,i}\,\sigma_s\,G_s^{n,i}\,ds \mid \mathcal{F}_i^n\Big)$$

We have $X_s^{c,n,i}\sigma_s = M_s^{n,i}\sigma_i^n + B_s^{n,i}\sigma_i^n + X_s^{c,n,i}(\sigma_s - \sigma_i^n)$, so the above is $2\sum_{l=1}^3 \mathbb{E}(H_k^n \mid \mathcal{F}_i^n)$, where

$$H_1^n = (\sigma_i^n)^2 \int_{T(n,i)}^{T'(n,i)} G_s^{n,i}\,ds \int_{T(n,i)}^s G_u^{n,i}\,dW_u$$
$$H_2^n = \sigma_i^n \int_{T(n,i)}^{T'(n,i)} G_s^{n,i}\,ds \int_{T(n,i)}^s G_u^{n,i}(\sigma_u - \sigma_i^n)\,dW_u$$
$$H_3^n = \int_{T(n,i)}^{T'(n,i)} \big(B_s^{n,i}\sigma_i^n + X_s^{c,n,i}(\sigma_s - \sigma_i^n)\big)\,G_s^{n,i}\,ds.$$

Using the definition of $G_s^{n,i}$, we can compute H_1^n explicitly in the same way as in Step 2 of the proof of Lemma A.9, to get $H_1^n = (\sigma_i^n)^2(H_n' + H_n'')$, where

$$H_n' = \sum_{r=1}^{h_n-1} (g_r^n)^2 \int_{T(n,i+r-1)}^{T(n,i+r)} (W_s - W_{T(n,i+r-1)})\,ds, \quad H_n'' = \sum_{l=1}^{h_n-2} g_l^n \Delta_{i+l}^n W \sum_{r=l+1}^{h_n-1} g_r^n \Delta(n, i+r),$$

and we make the following further decomposition $H_n'' = \sum_{m=1}^4 L_m^n$:

$$L_1^n = \sum_{1\leq l<r\leq h_n-1} g_l^n g_r^n \Delta_{i+l}^n W\Big(\Delta(n,i+r) - \tfrac{\Delta_n}{\alpha_{i+r-1}^n}\Big)$$
$$L_2^n = \Delta_n \sum_{1\leq l<r\leq h_n-1} g_l^n g_r^n \Delta_{i+l}^n W\Big(\tfrac{1}{\alpha_{i+r-1}^n} - \tfrac{1}{\alpha_{i+l}^n}\Big)$$
$$L_3^n = \Delta_n \sum_{1\leq l<r\leq h_n-1} g_l^n g_r^n \Delta_{i+l}^n W\Big(\tfrac{1}{\alpha_{i+l}^n} - \tfrac{1}{\alpha_{i+l-1}^n}\Big)$$
$$L_4^n = \Delta_n \sum_{1\leq l<r\leq h_n-1} g_l^n g_r^n \Delta_{i+l}^n W\,\tfrac{1}{\alpha_{i+l-1}^n}.$$

It is obvious that $\mathbb{E}(L_4^n \mid \mathcal{F}_i^n) = 0$, and by using (A.2), (A.6) with $V = 1/\alpha$, plus successive conditioning and the Cauchy–Schwarz inequality, one checks that $\mathbb{E}(|L_m^n| \mid \mathcal{F}_i^n)$ is smaller than $Kh_n^2\Delta_n^{3/2+\rho}$ when $m = 1$, than $Kh_n^2\Delta_n^{3/2}v_n$ when $m = 2$ and than $Kh_n^2\Delta_n^2$ when $m = 3$. With the same tools, we also get $\mathbb{E}(|H_n'| \mid \mathcal{F}_i^n) \leq Kh_n\Delta_n^{3/2} \leq Kv_n^2$, plus $\mathbb{E}(|H_k^n| \mid \mathcal{F}_i^n) \leq Kv_n^2$ for $k = 2,3$ (use here (A.17) when $k = 1$, and (A.6) for $V = \sigma$ when $k = 2$).

Putting all these together, we arrive at $\big|\mathbb{E}\big(z(1)_i^n\Delta(V,p)_j^n \mid \mathcal{F}_i^n\big)\big| \leq K(\Delta_n + \Delta_n^{1/2+\rho})$, hence $\big|\mathbb{E}\big(Z(1)_i^n\Delta(V,p)_j^n \mid \mathcal{F}_i^n\big)\big| \leq Kp(\sqrt{\Delta_n} + \Delta_n^\rho)$. This gives us (A.49) for $k = 1$, since $\rho > 1/4$ by hypothesis, and the proof is complete. □

At this stage, we obtain the first claim of Theorem 3.1 exactly as in Jacod et al. (2009): namely, by (A.13), (A.44) and Lemmas A.12, A.13, A.15 and A.16 the processes $V(p)^n = \Delta_n^{-1/4}(\widehat{C}_t^n - C_t - M(p)^n)$ satisfy

$$\lim_{p\to\infty}\ \limsup_{n\to\infty}\ \mathbb{P}\Big(|V(p)_t^n| > \varepsilon\Big) = 0$$

for all $\varepsilon > 0$ and all t. On the other hand, with a Brownian motion B independent of $\mathcal{F}$ being fixed and since with the notation (3.10) and (3.11) and those of Proposition A.17 we have $\beta(p)_s(\omega)^2 \to \beta_s(\omega)^2$ for all s, ω and $\beta(p)_s^2 \leq K$, we must have $Y(p)_t \stackrel{\mathbb{P}}{\longrightarrow} Y_t$. Then the first claim in Theorem 3.1 follows from Proposition A.17.

The last claim of the theorem follows from the first one and from (3.12) by well known properties of the $\mathcal{F}_\infty$-stable convergence, once noticed that the limit $\int_0^t \beta_s^2\,ds$ is $\mathcal{F}_t$-measurable.

It thus remains to prove (3.12), which amounts to the following three convergences:

$$\frac{\overline{h}_n}{\overline{h}_n'^2\sqrt{\Delta_n}}\,V_t^{n,1} \stackrel{\mathbb{P}}{\longrightarrow} 3\theta\phi(0)^2\int_0^t \frac{\sigma_s^4}{\alpha_s}\,ds, \tag{A.51}$$

$$\frac{1}{\overline{h}_n'\overline{h}_n\sqrt{\Delta_n}}\,V_t^{n,2} \stackrel{\mathbb{P}}{\longrightarrow} \frac{\phi(0)R}{\theta}\int_0^t \sigma_s^2\,\gamma_s^2\,ds, \tag{A.52}$$

$$\frac{1}{\overline{h}_n^3\sqrt{\Delta_n}}\,V_t^{n,3} \stackrel{\mathbb{P}}{\longrightarrow} \frac{R^2}{\theta^3}\int_0^t (\gamma_s)^4\alpha_s\,ds. \tag{A.53}$$

Proof of (A.51). In this proof we take $h_n = \overline{h}'_n$, so $V^{n,1} = \sum_{j=1}^{3} U^{n,j}$, where

$$U_t^{n,1} = \sum_{i=0}^{N_t^n - h_n} \zeta_i^n, \qquad U_t^{n,2} = \sum_{i=0}^{N_t^n - h_n} \zeta_i'^n, \qquad U_t^{n,3} = \sum_{i=0}^{N_t^n - h_n} \big((\widetilde{Y}_i^n)^4 \, 1_{\{|\widetilde{Y}_i^n| \le u'_n\}} - (\widetilde{Y}_i^{c,n})^4\big)$$
$$\zeta_i^n = \mathbb{E}\big((\widetilde{Y}_i^{c,n})^4 \mid \mathcal{K}_i^n\big), \qquad \zeta_i'^n = (\widetilde{Y}_i^{c,n})^4 - \zeta_i^n.$$

Then, the result will follow from the following three convergences:

$$\frac{\overline{h}_n}{\overline{h}'^2_n \sqrt{\Delta_n}} \, U_t^{n,1} \stackrel{\mathbb{P}}{\longrightarrow} 3\theta\phi(0)^2 \int_0^t \frac{\sigma_s^4}{\alpha_s}\, ds, \qquad j = 2, 3 \;\Rightarrow\; \frac{\overline{h}_n}{\overline{h}'^2_n \sqrt{\Delta_n}} \, U_t^{n,j} \stackrel{\mathbb{P}}{\longrightarrow} 0. \tag{A.54}$$

Lemma A.6 with $z = 2$ and $w = 1$ and $N_t^n \le Kt/\Delta_n$ yield $\mathbb{E}(|U_t^{n,3}|) \le K_t h_n'^{2+\eta} \Delta_n^{1+\eta}$ for some $\eta > 0$, hence (A.54) for $j = 3$.

Note that $\zeta_i'^n$ is $\mathcal{K}_{i+2h_n}^n$-measurable, and $\mathbb{E}((\zeta_i'^n)^2) \le K((h_n\Delta_n)^4 + 1/h_n^v)$. Then the same argument as for proving (S.11) gives us $\mathbb{E}((U_t^{n,2})^2) \le K_t(h_n^5\Delta_n^3 + 1/h_n^{v-1}\Delta_n)$. Recalling $h_n = \overline{h}'_n$, we deduce (A.54) for $j = 2$ by using (3.6).

For studying $U_t^{n,1}$, we first observe that $(\widetilde{Y}_i^{c,n})^4 = 6(\sigma_i^n)^4 \overline{G}_{0,0}^{n,i} + \sum_{k=1}^{7} \delta_i^{n,k}$, where

$$\delta_i^{n,1} = (\widetilde{X}_i^{c,n})^4 - 6(\sigma_i^n)^4 \overline{G}_{0,0}^{n,i}, \qquad \delta_i^{n,2} = 4\widetilde{X}_i^{c,n}\big((\widetilde{\varepsilon}_i^n)^3 - (\gamma_i^n)^3 \widetilde{\mathbf{r}}(0,0,0)_n\big)$$
$$\delta_i^{n,3} = 4\widetilde{X}_i^{c,n}(\gamma_i^n)^3 \widetilde{\mathbf{r}}(0,0,0)_n, \qquad \delta_i^{n,4} = 6(\widetilde{X}_i^{c,n})^2\big((\widetilde{\varepsilon}_i^n)^2 - (\gamma_i^n)^2 \widetilde{\mathbf{r}}(0,0)_n\big)$$
$$\delta_i^{n,5} = 6(\widetilde{X}_i^{c,n})^2 (\gamma_i^n)^2 \widetilde{\mathbf{r}}(0,0)_n, \qquad \delta_i^{n,6} = 4(\widetilde{X}_i^{c,n})^3 \widetilde{\varepsilon}_i^n, \qquad \delta_i^{n,7} = (\widetilde{\varepsilon}_i^n)^4.$$

We obtain $\mathbb{E}(|\mathbb{E}(\delta(k)_i^n \mid \mathcal{K}_i'^n)|) \le K\Delta_n$ by using (A.22) for $k = 1$ (since $(\overline{h}'_n\Delta_n)^{5/2} \le K\Delta_n$ because $\tau < 3/5$ in (3.6)), and (A.27) for $k = 7$, and (A.17) plus (A.23) for $k = 3$,5, and (A.17) plus the second part of (A.25) with $U = \widetilde{X}_i^{c,n}$ or $U = (\widetilde{X}_i^{c,n})^2$ for $k = 2$,4, and (A.26) with $U = (\widetilde{X}_i^{c,n})^3$ and again (A.17) for $k = 6$. Hence (A.54) for $j = 1$ amounts to having

$$\overline{U}_t^n := \frac{2\overline{h}_n}{\overline{h}'^2_n \sqrt{\Delta_n}} \sum_{i=0}^{N_t^n - \overline{h}'_n} (\sigma_i^n)^4 \overline{G}_{0,0}^{n,i} \stackrel{\mathbb{P}}{\longrightarrow} \theta\phi(0)^2 \int_0^t \frac{\sigma_s^4}{\alpha_s}\, ds. \tag{A.55}$$

This is a simple consequence of (A.39) and Lemma A.11, both with $h_n = \overline{h}'_n$, plus (3.6). □

Proof of (A.52). We take again $h_n = \overline{h}'_n$ in this proof. Set

$$U_t^n = \sum_{i=1}^{N_t^n - h_n - 6k_n} (\gamma_i^n)^2 (\widetilde{Y}_i^{c,n})^2, \qquad V'(m)_t^n = V(m)_t^n - r(m) U_t^n.$$

(2.8) yields $\big|R - \sum_{m=-k'_n}^{k'_n} r(|m|)\big| \le K/k_n'^{v-1}$, hence by the definition of $V_t^{n,2}$ we have

$$\big|\, V_t^{n,2} - RU_t^n \big| \le \frac{K}{k_n'^{v}}\, U_t^n + A_t^n, \qquad A_t^n = \sum_{m=-k'_n}^{k'_n} \big| r(m)\, V'(|m|)_t^n \big|.$$

(A.34) with $h_n = \overline{h}_n$ and $\sum_m |r(m)| < \infty$ imply $\mathbb{E}(A_t^n) \le K h_n^{1+\eta} \Delta_n^{\eta}$ for some $\eta > 0$, hence $\frac{1}{\overline{h}_n \overline{h}'_n \sqrt{\Delta_n}} A_t^n \stackrel{\mathbb{P}}{\longrightarrow} 0$ (recall $\overline{h}_n \sqrt{\Delta_n} \to \theta$), and it remains to prove that

$$\frac{1}{h_n}\, U_t^n \stackrel{\mathbb{P}}{\longrightarrow} \phi(0) \int_0^t \sigma_s^2\, \gamma_s^2\, ds. \tag{A.56}$$

For this, we reproduce in a simpler way the previous proof. We have $U^n = U_t'^n + U_t''^n$, where

$$U_t'^n = \sum_{i=0}^{N_t^n - h_n - 6k_n} \zeta_i'^n, \qquad U_t''^n = \sum_{i=0}^{N_t^n - h_n - 6k_n} \zeta_i''^n$$
$$\zeta_i^n = (\gamma_i^n)^2\, \mathbb{E}\big((\widetilde{Y}_i^{c,n})^2 \mid \mathcal{K}_i^n\big), \qquad \zeta_i'^n = (\gamma_i^n)^2 (\widetilde{Y}_i^{c,n})^2 - \zeta_i^n,$$

so (A.56) will result from the following two convergences:

$$\frac{1}{h_n}\, U_t'^n \stackrel{\mathbb{P}}{\longrightarrow} \phi(0) \int_0^t \sigma_s^2\, \gamma_s^2\, ds, \qquad \frac{1}{h_n}\, U_t''^n \stackrel{\mathbb{P}}{\longrightarrow} 0. \tag{A.57}$$

Since $\zeta_i''^n$ is $\mathcal{K}_{i+2h_n}^n$-measurable and $\mathbb{E}((\zeta_i'^n)^2) \le K(h_n\Delta_n)^2$, as in the previous proof we deduce $\mathbb{E}((U_t^{n,2})^2) \le K_t h_n^3 \Delta_n$, hence the second part of (A.57). For the first part, we observe that $(\gamma_i^n \widetilde{Y}_i^{c,n})^2 = (\gamma_i^n)^2 \widehat{c}_i^n + \sum_{k=1}^{3} \delta_i^{n,k}$, where

$$\delta_i^{n,1} = (\gamma_i^n)^2 \widehat{X}_i^{c,n}, \qquad \delta_i^{n,2} = 2(\gamma_i^n)^2 \widetilde{X}_i^{c,n} \widetilde{\varepsilon}_i^n, \qquad \delta_i^{n,3} = (\gamma_i^n)^2 (\widetilde{\varepsilon}_i^n)^2.$$

Then $\mathbb{E}(|\mathbb{E}(\delta(k)_i^n \mid \mathcal{K}_i'^n)|) \leq K h_n \Delta_n^{5/4}$, by using (A.19) for $k=1$, and (A.27) for $k=3$, and (A.17) plus (A.26) with $U=(\gamma_i^n)^2 \widetilde{X}_i^{c,n}$ for $k=2$. Then, if we also use (A.21), (A.38) and Lemma A.11, we readily deduce the first part of (A.57). This completes the proof. $\square$

Proof of (A.53). First, Lemma A.11 yields

$$U_t^n := \Delta_n \sum_{i=0}^{N_t^n - 11k_n} (\gamma_i^n)^4 \ \xrightarrow{\mathbb{P}}\ \int_0^t (\gamma_s)^4 \alpha_s \, ds.$$

Second, $\left|R^2 - \sum_{m,m'=-k_n'}^{k_n'} r(|m|) r(|m'|)\right| \leq K / k_n'^{v-1}$ implies

$$\left|V_t^{n,3} - \tfrac{R^2}{\Delta_n} U_t^n\right| \leq \tfrac{K}{k_n'^{v} \Delta_n} U_t^n + A_t^n, \qquad \text{where}$$

$$A_t^n = \sum_{m,m'=-k_n'}^{k_n'} \left| r(m) r(m') \left(\overline{U}(|m|,|m'|)_t^n - \tfrac{r(m)r(m')}{\Delta_n} U_t^n\right)\right.$$

Then (A.53) is an obvious consequence of (A.34) and (3.6), plus $\sum_m |r(m)| < \infty$. $\square$

Appendix B. Supplementary data

Supplementary material Proofs of Lemmas can be found online at https://doi.org/10.1016/j.jeconom.2018.09.006.

References

Aït-Sahalia, Y., Fan, J., Li, Y., 2013. The leverage effect puzzle: Disentangling sources of bias at high frequency. J. Financ. Econ. 109, 224–249.
Aït-Sahalia, Y., Jacod, J., Li, J., 2012. Testing for jumps in noisy high frequency data. J. Econometrics 168, 207–222.
Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. J. Econometrics 160, 190–203.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. J. Econometrics 162, 149–169.
Christensen, K., Podolskij, M., Vetter, M., 2013. On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. J. Multivariate Anal. 120, 59–84.
Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econom. Statist. 24, 127–161.
Hautsch, N., Podolskij, M., 2013. Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: theory, implementation, and empirical evidence. J. Bus. Econom. Statist. 31, 165–183.
Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bonds and currency options. Rev. Financ. Stud. 6, 327–343.
Ikeda, S., 2015. Two-scale realized kernels: a univariate case. J. Financ. Econ. 13, 126–165.
Ikeda, S., 2016. A bias-corrected estimator of the covariation matrix of multiple security prices when both microstructure effects and sampling durations are persistent and endogenous. J. Econometrics 193, 203–214.
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Process. Appl. 119, 2249–2276.
Jacod, J., Li, L., Zheng, X., 2017. Statistical properties of microstructure noise. Econometrica 85, 1133–1174.
Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes, second ed. Springer-Verlag, Berlin.
Jing, B., Liu, Z., Kong, X., 2014. On the estimation of integrated volatility with jumps and microstructure noise. J. Bus. Econom. Statist. 32, 457–467.
Kalnina, I., Linton, O., 2008. Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. J. Econometrics 147, 47–69.
Li, Y., Mykland, P.A., Renault, E., Zhang, L., Zheng, X., 2014. Realized volatility when sampling times are possibly endogenous. Econometric Theory 30, 580–605.
Li, Y., Xie, S., Zheng, X., 2016. Efficient estimation of integrated volatility incorporating trading information. J. Econometrics 195, 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, 187–222.
Li, Y., Zhang, Z., Zheng, X., 2013. Volatility inference in the presence of both endogenous time and microstructure noise. Stochastic Process. Appl. 123, 2696–2727.
Mancini, C., 2001. Disentangling the jumps of the diffusion in a geometric jumping brownian Motion. Giornale dell'Istituto Italiano degli Attuari 64, 19–47.
Tsay, R., 2002. Analysis of Financial Time Series. John Wiley & Sons, Inc, Hoboken, NJ.
Ukabata, M., Oya, K., 2009. Estimation and testing for dependence in market microstructure noise. J. Financial Econometrics 7, 106–151.
Xia, N., Zheng, X., 2018. On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations. Ann. Statist. 46, 500–525.
Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159, 235–250.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12, 1019–1043.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100, 1394–1411.
Zheng, X., Li, Y., 2011. On the estimation of integrated covariance matrices of high dimensional diffusion processes. Ann. Statist. 39, 3121–3151.

Further Reading

Jacod, J., Protter, P., 2012. Discretization of Processes. Springer-Verlag, Berlin.
Kalnina, I., 2011. Subsampling high frequency data. J. EconometricsâĂŹ 161, 262–283.
Podolskij, M., Vetter, M., 2009. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli 15, 634–658.
Varneskov, R.T., 2016. Flat-top realized kernel estimation of quadratic covariation with nonsynchronous and noisy asset prices. J. Bus. Econom. Statist. 34, 1–22.