Finance Stoch (2015) 19:261–293
DOI 10.1007/s00780-015-0255-1

Spot volatility estimation using delta sequences

Cecilia Mancini[1] · **Vanessa Mattiussi**[2] ·
Roberto Renò[3]

Received: 21 October 2009 / Accepted: 7 June 2013 / Published online: 20 February 2015

Abstract We introduce a unifying class of nonparametric spot volatility estimators based on delta sequences and conceived to include many of the existing estimators in the field as special cases. The full limit theory is first derived when unevenly sampled observations under infill asymptotics and fixed time horizon are considered, and the state variable is assumed to follow a Brownian semimartingale. We then extend our class of estimators to include Poisson jumps or financial microstructure noise in the observed price process. This work makes different approaches (kernels, wavelets, Fourier) comparable. For example, we explicitly illustrate some drawbacks of the Fourier estimator. Specific delta sequences are applied to data from the S&P 500 stock index futures market.

Keywords Spot volatility · High-frequency data · Microstructure noise · Dirac delta · Fourier estimator

Mathematics Subject Classification (2010) 91G70

JEL Classification C13 · C14 · C22 · G1

Electronic supplementary material The online version of this article (doi:10.1007/s00780-015-0255-1) contains supplementary material, which is available to authorized users.

✉ C. Mancini
cecilia.mancini@unifi.it

V. Mattiussi
vanessa.mattiussi.1@city.ac.uk

R. Renò
reno@unisi.it

1 Università di Firenze, via delle Pandette 9, 50127 Firenze, Italy

2 City University, Northampton Square, London EC1V 0HB, UK

3 Università di Siena, Piazza S. Francesco 7, 53100 Siena, Italy

1 Introduction

In the last decade, the larger availability of high-frequency financial data sets has spawned considerable econometric research on integrated volatility and in particular on realized volatility (reviews on the topic can be found in Barndorff-Nielsen and Shephard [11], Bandi and Russell [9], and Mancini and Calvori [38]). However, interest has recently moved to spot volatility, both in theory and in practice. For example, Jacod and Rosenbaum [28] show that efficient estimation of integrated volatility functionals can be achieved by using Riemann-like integration of preliminary spot volatility estimates. Spot volatility is also relevant in other financial applications and in particular in detecting the fine structure of price dynamics, as for testing for the presence of jumps [1, 12], or of cojumps [30], or for detecting the correlation between price and volatility jumps [31, 8]. Spot volatility is also the crucial ingredient in option pricing with stochastic volatility, where the initial volatility value, in addition to the initial value of the underlying, is needed to price the option.

Given the growing importance of spot volatility in financial applications, and given that many types of estimators have already been proposed, we study a wide and unifying class of estimators for the spot volatility of a univariate semimartingale. Importantly, we allow jumps in the volatility process (in this case, our estimator converges to the average of spot volatilities observed immediately before and after the volatility jump) and, with suitable adjustments, for the presence of Poisson jumps or microstructure noise in the price process.

The literature on spot volatility estimation is large. One way of estimating instantaneous volatility consists in assuming that the coefficient is a deterministic function of the observable state variable X, and nonparametric techniques can be applied both in the absence (see Florens-Zmirou [18], Bandi and Phillips [6], Renò [49], and Hoffmann [22]) and in the presence of jumps in X (see Johannes [32], Bandi and Nguyen [5], and Mancini and Renò [39]). Fully nonparametric methods, when volatility is instead a càdlàg process, have been studied by Malliavin and Mancino [35, 36] and Kristensen [34] in the absence of jumps, and by Zu and Boswijk [56], Hoffmann et al. [21], and Ogawa and Sanfelici [45] in the absence of jumps but with noisy observations. Related studies include the idea of rolling sample volatility estimators in Foster and Nelson [19], see also Andreou and Ghysels [4]; the theory of spot volatility estimation developed in Bandi and Renò [7]; and the kernel-based methods of Fan and Wang [16] and Mykland and Zhang [43]. In the presence of jumps in price (but absence of noise), spot volatility estimation has been studied by Jacod and Protter [27], Ngo and Ogawa [44], Aït-Sahalia and Jacod [1], and Dobrev et al. [14]. Further alternatives are studied in Alvarez et al. [2], Genon-Catalot et al. [20], and Hoffmann [23].

Using kernels (as in Kristensen [34]) is one of the most popular methods. However, some volatility estimators in the above-mentioned literature are not of kernel type. The purpose of our study is then to propose a unifying approach. This leads to the specification of a large class of nonparametric estimators of instantaneous volatility, which includes most of the aforementioned methods and thus allows a comparison between them. Our idea is to write a spot volatility estimator as the convolution of squared price returns with a sequence of functions, known as *delta sequence*, which converges to a Dirac delta function concentrating all the mass around one point in time (for applications of delta sequences in statistics, see for instance Watson and Leadbetter [53], and Walter and Blum [52]). We provide a theory supporting this intuition. In particular, we extend the kernel estimator of Kristensen [34] by proving that a traditional kernel function can be seen as a delta sequence. Our class is shown to be reasonably wide, and it includes the Fejér sequence used in the work of Malliavin and Mancino [36], the indicator function (a kernel itself) used in the work of Jacod and Rosenbaum [28] and in Chap. 13 of Jacod and Protter [27], and linear wavelets as in Walter [51].

The study of the asymptotic theory (see Sect. 2) reveals that the estimators within the class are, under suitable conditions, mixed normally distributed when the number of (not necessarily equally spaced) observations diverges to infinity in a fixed interval $[0, T]$ and the maximum duration between the observations shrinks to zero. Our findings are derived under mild assumptions on the coefficients of the stochastic differential equation for X. In Sect. 3, we allow microstructure noise in the data and use a two-scale volatility technique, similar to the one in Zhang et al. [55], to make our estimator robust against the noise. In addition, in the presence of finite activity jumps in X, we threshold the returns as in Mancini [37] to filter out jumps from the observed price process. In Sect. 4, we show that the Fourier estimator of spot volatility proposed in Malliavin and Mancino [36] can be written as the sum of an estimator constructed with the Fejér delta sequence plus a cross product term, thus showing that in the absence of noise (such as when sampling sparsely), the Fourier estimator is less appealing in terms of both estimation error variance and computational burden. Section 5 presents an empirical application to high-frequency stock index futures, where the above estimators are applied to detect intraday volatility dynamics. Section 6 concludes. All the proofs are contained in Appendix A, with the exception of those of Lemma A.1 and Theorem 3.3, which are contained in the online available Appendix B of this paper.

2 Spot volatility estimation in the basic setting

In what follows, we consider a univariate logarithmic price process (X_t) defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathrm{P})$ satisfying the usual conditions. Our results are based on the set of assumptions outlined below.

Assumption 2.1 (i) The logarithmic price X is the solution of the stochastic differential equation

$$dX_t = \mu_t\,dt + \sigma_t\,dW_t,$$

where the initial condition X_0 is measurable with respect to $\mathcal{F}_0$, W is a standard Brownian motion defined on the filtered probability space, and μ, σ are adapted processes with càdlàg paths.

(ii) Given a fixed point $\bar{t} \in [0, T]$, let $B_\varepsilon(\bar{t}) = [\bar{t} - \varepsilon, \bar{t} + \varepsilon]$ with fixed $\varepsilon > 0$, and assume that there exists a constant Γ, a sequence of stopping times $\tau_m \uparrow \infty$, and constants $C_{\bar{t}}^{(m)}$ such that for all m, for $(\omega, s) \in (\Omega \times B_\varepsilon(\bar{t})) \cap [\![0, \tau_m[\![$ and $u \in B_\varepsilon(\bar{t})$, we have

$$\mathrm{E}_{u\wedge s}[|\sigma_u - \sigma_s|^2] \leq C_{\bar{t}}^{(m)} |u - s|^\Gamma, \tag{2.1}$$

where $\mathrm{E}_t[\cdot]$ denotes $\mathrm{E}[\cdot|\mathcal{F}_t]$.

The class of processes for (σ_t) we wish to estimate pointwise is larger than the class of processes with differentiable paths, and it includes the important case where σ is itself generated by a Brownian motion as in a stochastic volatility model. Note that every Itô semimartingale of the following type is càdlàg and satisfies (2.1) (with $\Gamma = 1$):

$$\sigma_t = \int_0^t \tilde{\mu}_s\, ds + \int_0^t \tilde{\sigma}_s\, d\tilde{W}_s + \int_0^t \int_{x:|\tilde{\gamma}(\omega,x,s)|>1} \tilde{\gamma}(\omega, x, s)\mu(dx, ds)$$
$$+ \int_0^t \int_{x:|\tilde{\gamma}(\omega,x,s)|\leq 1} \tilde{\gamma}(\omega, x, s)\big(\mu(dx, ds) - \nu(dx)\big)\, ds,$$

where μ is a Poisson measure counting the jumps of σ, g is a deterministic strictly positive function such that $\int (g^2(x) \wedge 1)\nu(dx) < \infty$, and $\tilde{\mu}, \tilde{\sigma}, \sup_x \frac{|\tilde{\gamma}(\omega,x,s)|}{g(x)}$ are predictable processes. Property (2.1) is a consequence of the Burkholder–Davis–Gundy (BDG) inequality. In particular, the estimator we are going to propose is robust to jumps in volatility.

In order to work with irregular sampling, we adapt to our setting the concept of quadratic variation of time defined in Mykland and Zhang [42].

Assumption 2.2 The process X is observed $n + 1$ times at deterministic instants $0 = t_0 < t_1 < \cdots < t_n = T$, not necessarily equally spaced and with T fixed. We set $\Delta_i = t_i - t_{i-1}$ and $\overline{\Delta}_n = \frac{T}{n}$ and assume that $\max_{i=1,\ldots,n} \Delta_i = O(\overline{\Delta}_n)$. The quadratic variation of time up to a given $t \leq T$ is defined as $H(t) = \lim_{n\to\infty} H_n(t)$, where

$$H_n(t) = \frac{1}{\overline{\Delta}_n} \sum_{t_i \leq t} (\Delta_i)^2 .$$

Assuming that the above limit exists, we require that H is Lebesgue-almost surely differentiable in $[0, T]$, with H' bounded such that for some $K \geq 0$ (not depending on i),

$$\left| H'(t_i) - \frac{\Delta_i}{\overline{\Delta}_n} \right| \leq K\Delta_i \quad \text{for any } t_i \text{ in which } H \text{ is differentiable.} \tag{2.2}$$

In the special case of equally spaced observations, $\Delta_i = \overline{\Delta}_n$, $H'(t) = 1$, and (2.2) is satisfied with $K = 0$. When the observations are more (less) concentrated around t, we have $H'(t) < 1$ $(H'(t) > 1)$. The assumption $\max_{i=1,\ldots,n} \Delta_i = O(\overline{\Delta}_n)$ is technical and means that the partition should not vary asymptotically too wildly with respect to the equally spaced partition.

The condition in (2.2) for the partition $\{t_i\}_i$ is different from condition (v) in assumption A of [42]. For instance, consider the sequence of partitions where the amplitude of the first $[n/2]$ intervals $(t_{i-1}, t_i]$ is 2Δ and of the remaining $n-[n/2]$ is Δ. Then $\Delta = \frac{\bar{\Delta}}{1+[\frac{n}{2}]\frac{1}{n}}$ and $H(t) = 4t/3\ I_{\{t\le T_1\}} + (4T/9 + 2t/3)I_{\{t>T_1\}}$ with $T_1 = 2T/3$. This function H is not differentiable in T_1; however, we have that $|H'(t_i) - \frac{\Delta_i}{\bar{\Delta}_n}| \le \frac{2}{T}\Delta_i$, $\forall n$ for all other points. It follows that our Assumption 2.2 is satisfied. On the contrary, if we only consider the time points t where H is differentiable, the assumption of Mykland and Zhang [42] is not fulfilled since

$$\sup_{t>T_1}\left|\frac{H_n(t) - H_n(t-\sqrt{\bar{\Delta}_n})}{\sqrt{\bar{\Delta}_n}} - H'(t)\right| \longrightarrow +\infty.$$

Denote the (log-price) return by $\Delta X_i = X_{t_i} - X_{t_{i-1}}$. Our proposed estimator takes the form of a discrete convolution

$$\widehat{\sigma}^2_{n,f}(\bar{t}) = \sum_{i=1}^{n} f_n(t_{i-1} - \bar{t})\,(\Delta X_i)^2, \quad \bar{t} \in D, \tag{2.3}$$

where $(f_n(\cdot))$ is a given sequence of real functions belonging to the class specified below.

Definition 2.3 A sequence $F = \{f_n, n \in \mathbb{N}\}$ of functions $f_n : D \to \mathbb{R}$, with $D \subseteq \mathbb{R}$ a given set and $0 \in \mathring{D}$, where $\mathring{D}$ indicates the interior of D, is said to be a *delta sequence* if for all processes (σ_t) satisfying Assumption 2.1, as $n \to \infty$,

$$\int_0^T f_n(s-\bar{t})\sigma_s^2\,ds = (\sigma^2)^\star_{\bar{t}} + R_n^{(\sigma^2)}(\bar{t}), \tag{2.4}$$

$$\frac{1}{f_n(0)}\int_0^T f_n^2(s-\bar{t})\sigma_s^2\,ds = c_f(\sigma^2)^\star_{\bar{t}} + o_p(1), \tag{2.5}$$

$$\frac{1}{f_n^2(0)}\int_0^T f_n^4(s-\bar{t})\sigma_s^2\,ds = O_p\big(f_n(0)\big), \tag{2.6}$$

where $R_n^{(\sigma^2)}(\bar{t}) = o_p(1)$ and

$$(\sigma^2)^\star_t = \big(\psi_f^+\sigma_t^2 + \psi_f^-\sigma_{t-}^2\big)I_{\{t\in(0,T)\}} + \psi_f^-\sigma_{T-}^2 I_{\{t=T\}} + \psi_f^+\sigma_0^2 I_{\{t=0\}},$$

where $\int_{x<0} f_n(x)\,dx \to \psi_f^-$ and $\psi_f^+ = 1 - \psi_f^-$ ($\psi_f^- = \psi_f^+ = \frac{1}{2}$ for symmetric delta sequences).

Throughout the paper, it is intended that the integrals are defined over the intersection with $s \in D$. The notations O_p, o_p, and $O_{a.s}$ are explained in Appendix A.

Note that if σ^2 is continuous in $\bar{t} \in (0,T)$, then $(\sigma^2)^\star_{\bar{t}} = \sigma^2_{\bar{t}}$. If we estimate instead at the boundaries ($\bar{t} = 0$ or $\bar{t} = T$), then we have to weight for the exact mass of the delta sequence respectively at the right and at the left of $\bar{t}$.

Condition (2.4) resembles the typical definition of a delta sequence in analysis. The technical conditions (2.5) and (2.6) are required to guarantee the central limit theorem. Delta sequences have been introduced in statistics to estimate the density of a random variable; see, for example, Watson and Leadbetter [53]. The main result of this section is stated under a set of additional conditions that we collect in the

Assumption 2.4 We assume that $F = \{f_n, n \in \mathbb{N}\}$ is a delta sequence with $f_n(0) \to +\infty$ and $\int_D f_n(x)\,dx \to 1$ and that further the functions f_n satisfy

(i) $\sup_{x\in D} |f_n(x)| \le C f_n(0)$ for a suitable constant C.
(ii) f_n is Lipschitz in a neighborhood of 0 with a Lipschitz constant L_n such that $L_n\sqrt{\overline{\Delta}_n/f_n(0)} \to 0$; further, either $f_n \ge 0$ or $\overline{\Delta}_n^{\Gamma/2} \sum_i |f_n(t_{i-1} - \bar{t})\Delta_i| \to 0$.
(iii) There exists a constant $M_\varepsilon > 0$ not depending on n for which

$$\sup_{x\in B_\varepsilon^c(0)} |f_n(x)| \le M_\varepsilon. \tag{2.7}$$

Theorem 2.7 below holds also when the condition $L_n\sqrt{\overline{\Delta}_n/f_n(0)} \to 0$ in Assumption 2.4(ii) is replaced with the (less stringent but less direct to be verified) condition that $\sum_{i=1}^n \int_{t_{i-1}}^{t_i} |f_n(s-\bar{t}) - f_n(t_{i-1}-\bar{t})|\,ds \to 0$ for the consistency part, and with $f_n(0)\bar{\Delta} \to 0$ and $\sum_{i=1}^n \int_{t_{i-1}}^{t_i} |f_n(s-\bar{t}) - f_n(t_{i-1}-\bar{t})|\,ds/\sqrt{\overline{\Delta}_n f_n(0)} \to 0$ for the CLT part. Definition 2.3 is not straightforward to be verified for a given sequence (f_n). For this reason, we specify the following proposition, which involves a set of sufficient conditions using only the features of (f_n) instead of involving also the features of the process (σ_t).

Proposition 2.5 *Consider a sequence of nonnegative functions $f_n : D \to \mathbb{R}$, with $D \subset \mathbb{R}$ and $0 \in \mathring{D}$, such that as $n \to \infty$, conditions* (i)–(iii) *in Assumption 2.4 are fulfilled, and furthermore*

(iv)

$$\int_D f_n(x)\,dx \longrightarrow 1, \tag{2.8}$$

(v) *there exists a sequence $\varepsilon_n \to 0$ such that*

$$\int_{-\varepsilon_n}^{\varepsilon_n} f_n(x)\,dx \longrightarrow 1, \tag{2.9}$$

(vi)

$$\int_D \frac{f_n^2(x)}{f_n(0)}\,dx \longrightarrow c_f, \tag{2.10}$$

where c_f is a real constant.

Then (f_n) is a delta sequence.

In (2.5) we have chosen to normalize $f_n^2(x)$ by $f_n(0)$, but alternatively $(f_n(0))$ can be replaced by any sequence (a_n) able to deliver similar results, such as $a_n = \int f_n^2(x)\,dx$.

Some relevant examples of sequences (f_n) satisfying Assumption 2.4 are listed below. Other examples can be derived from Walter and Blum [52].

Example 1 (Kernels) Kernel estimators, used by Kristensen [34] to estimate spot volatility, can indeed be used to generate a class of delta sequences. Consider a function $K : \mathbb{R} \to \mathbb{R}$ and a positive sequence $h_n \to 0$, and define

$$f_n(x) = \frac{1}{h_n} K\left(\frac{x}{h_n}\right). \tag{2.11}$$

The sequence (h_n) is typically called *bandwidth*, and since $f_n(0) = \frac{1}{h_n} K(0)$, we can interpret $f_n(0)$ as the inverse of the bandwidth. In the case in which we write the delta sequence as (2.11), Assumption 2.4 can be reformulated as follows.

Assumption 2.6 (Assumption 2.4 for kernels)

(1) $\int_{-\infty}^{+\infty} K(x)\,dx = 1$ and $\int_{-\infty}^{\infty} K^2(x)\,dx = c_2$ $(c_f = \frac{c_2}{K(0)})$.
(2) $\sup_{x\in\mathbb{R}} |K(x)| \le C K(0)$.
(3) K is almost everywhere differentiable, and K' is bounded.
(4) h_n is such that $\sup_{x\in\mathbb{R}} |K'(x h_n^{-1})| \sqrt{\Delta_n / h_n^3} \to 0$.
(5) $\sup_{x\in B_\varepsilon^c(\bar{t})} |\frac{1}{h_n} K(\frac{x}{h_n})| \le M_\varepsilon$.

For example, the *Gaussian kernel*

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

has $c_2 = \frac{1}{2\sqrt{\pi}}$ and $c_f = \frac{1}{\sqrt{2}}$, and Assumption 2.6 is readily verified, whereas the *Epanechnikov kernel*

$$K(x) = \frac{3}{4}(1 - x^2) I_{\{|x|\le 1\}}$$

has $c_2 = \frac{3}{5}$ and $c_f = \frac{4}{5}$ and also verifies Assumption 2.6. The *indicator kernel*

$$K(x) = \frac{1}{2} I_{\{|x|\le 1\}}$$

also verifies Assumption 2.6 and has $c_2 = \frac{1}{2}$ and $c_f = 1$.

Example 2 (Trigonometric functions) Trigonometric functions used in Fourier analysis are traditional approximants of the Dirac delta and naturally appear in the construction of the Fourier estimator of Malliavin and Mancino [35]. The first example is the Dirichlet sequence given by $g_n(x) = \frac{1}{2\pi} D_{N_n}(x)$, with domain $[-\pi, \pi]$, where

$$D_N(x) := \sum_{|h|\le N} e^{ihx} = \frac{\sin((N + \frac{1}{2})x)}{\sin\frac{x}{2}}, \tag{2.12}$$

and (N_n) is a diverging sequence. The Dirichlet sequence can be negative at some points. A positive trigonometric example, which will become crucial in Sect. 4, is given by $f_n(x) = \frac{1}{2\pi} F_{N_n}(x)$ with domain $(-\pi, \pi)$, where (F_N) is the Fejér sequence

$$F_N(x) := \sum_{|s|\le N} \left(1 - \frac{|s|}{N+1}\right) e^{\mathrm{i}sx} = \frac{1}{N+1} \left(\frac{\sin(\frac{N+1}{2}x)}{\sin\frac{x}{2}}\right)^2, \tag{2.13}$$

and (N_n) is another diverging sequence. The following properties hold for all N:

$$\begin{aligned} &\frac{1}{2\pi}\int_{-\pi}^{\pi} F_N(x)\,dx = 1, \qquad && \frac{1}{F_N(0)}\int_{-\pi}^{\pi} F_N^2(x)\,dx = \frac{4\pi}{3}, \\ &\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(x)\,dx = 1, \qquad && D_N^2(x) = (2N+1)F_{2N}(x), \end{aligned}$$

proving that f_n and g_n integrate to 1 and that $c_f = \frac{2}{3}$ and $c_g = 1$. Now notice that $1/|\sin(x/2)| \le 1/\sin(\varepsilon/2)$ if $\varepsilon \le |x| \le \pi$ with $0 < \varepsilon < \pi$. This easily proves conditions (iv) and (vi) in Proposition 2.5 for f_n. Moreover, with $\varepsilon = \varepsilon_n \to 0$, we have

$$\int_{\varepsilon_n \le |x| \le \pi} F_{N_n}(x)\,dx \le \frac{1}{N_n+1}\frac{1}{\sin^2(\varepsilon_n/2)} 2(\pi - \varepsilon_n),$$

which converges to zero if $\varepsilon_n^2 N_n \to \infty$. This proves, together with the remaining trivial conditions in Proposition 2.5, that (f_n) is a delta sequence. However, note that the Fejér delta sequence (f_n) is not of kernel type.

The following theorem derives the asymptotic distribution of the proposed volatility estimator (2.3). We use $\mathbf{MN}(0, V)$ to denote a mixed normal distribution with stochastic variance V.

Theorem 2.7 *Let Assumptions 2.1, 2.2, 2.4 hold. As $n \to \infty$, let $f_n(0) \to \infty$ in such a way that $f_n(0)\overline{\Delta}_n \to 0$. Then for any $\bar{t} \in [0, T]$, we have*

$$\widehat{\sigma}^2_{n,f}(\bar{t}) \xrightarrow{\mathrm{P}} (\sigma^2)^\star_{\bar{t}}.$$

If furthermore $R_n^{(\sigma^2)}(\bar{t}) = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n}\right)$, then

$$\frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}} \left(\widehat{\sigma}^2_{n,f}(\bar{t}) - (\sigma^2)^\star_{\bar{t}}\right) \xrightarrow{\mathcal{L}-(s)} \mathbf{MN}\Big(0, 2c_f\big(H'\sigma^4\big)^\star_{\bar{t}}\Big),$$

where the convergence is stable in law.

A similar result has been obtained in Kristensen [34] when $f_n(x)$ is of the form (2.11). On the notion of stable convergence, see Jacod [24].

Remark 2.8 (On the validity of a CLT) The crucial condition for the validity of our CLT is $R_n^{(\sigma^2)}(\bar{t}) = o_p(\sqrt{f_n(0)\overline{\Delta}_n})$. This condition is, however, typically satisfied with suitable choices of the sequence $(f_n(0))$ (or, in the kernel case, of the bandwidth). In Appendix A, we explicitly prove that for the Gaussian, Epanechnikov, and indicator kernels, this condition is fulfilled when $\Gamma < 6$ and $nh_n^{\Gamma+1} \to 0$, and for other kernels the condition can be verified in a similar way. The Fejér sequence is explicitly treated in Proposition 4.2.

Remark 2.9 (Small sample correction) In small samples, it is advisable to use the estimator

$$\widehat{\widehat{\sigma}}^2_{n,f}(\bar{t}) = \frac{\sum_{i=1}^n f_n(t_{i-1} - \bar{t})(\Delta X_i)^2}{\sum_{i=1}^n f_n(t_{i-1} - \bar{t})\Delta_i}, \tag{2.14}$$

for which it is immediate to derive the same asymptotic results as in Theorem 2.7 given that

$$\sum_{i=1}^n f_n(t_{i-1} - \bar{t})\Delta_i \longrightarrow 1 \quad \text{as } n \to \infty.$$

Remark 2.10 (Choice of the optimal f_n) The choice of the optimal sequence (f_n) relies on the usual bias-variance tradeoff considerations (see, for example, Fan and Yao [17] or the discussion in Kristensen [34]). From the proof of Theorem 2.7 we can see that the bias depends both on the choice of the kernel and the regularity of σ. For example, in the case in which f_n is the indicator function, we get (see the proof of Remark 2.8) that the bias is $O(f_n(0)^{-\Gamma/2})$, and, given that the variance is $O(f_n(0)\overline{\Delta}_n)$, we get that the optimal choice of $f_n(0)$ is proportional to $(\overline{\Delta}_n)^{-\frac{1}{1+\Gamma}}$, and the rate of convergence of the spot volatility estimator is $n^{1/4}$. For the choice of the optimal delta sequence, van Eeden [15] suggests the usage of a double exponential kernel, that is, $f_n(x) = f_n(0)e^{-\frac{2|x|}{f_n(0)}}$. To get some insight on this problem, we simulate the model

$$\begin{aligned} dX_t &= \mu + \sigma_t \, dW_t^{(1)}, \\ d(\log \sigma_t^2) &= \eta \, dW_t^{(2)} + dJ_t, \end{aligned} \tag{2.15}$$

where corr $(dW^{(1)}, dW^{(2)}) = \rho \, dt$, and the jump occurs exactly in $\bar{t} = 0.5$, with size normally distributed with mean 1.44 and standard deviation 0.11. We set $\eta = 0.6$, $\rho = -0.25$, and $\mu = 0.06/252$ (these parameters are based on estimates in [8]). We generate 1,000 paths, and we set $n = 1{,}000$ and $f_n(0) = 10$ and consider six delta sequences: indicator, triangular, Gaussian, Epanechnikov, Fejér, and double exponential. We adopt the correction (2.14). The (relative) root mean square error in estimating $(\sigma^2)^\star_{\bar{t}=0.5}$ is shown in Fig. 1, showing that there is no substantial difference between the delta sequences, even if the double exponential kernel seems to present some advantages. On the contrary, the choice of $f_n(0)$ (not discussed here) is crucial.

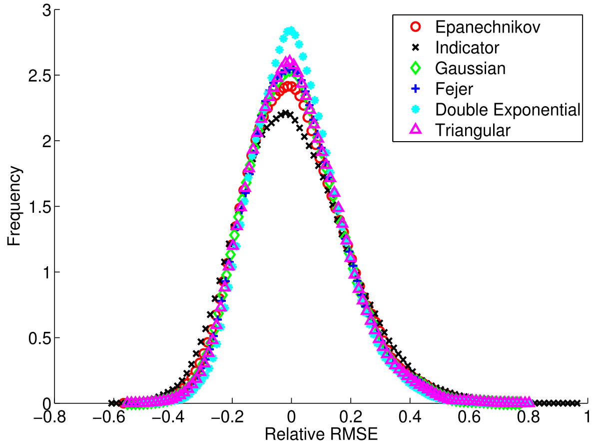

Fig. 1 Relative RMSE distribution for the estimation of $(\sigma^2)^\star_{\bar{t}}$ on 1,000 replications of model (2.15) for the six different delta sequences listed in the legend

3 Estimation in the presence of microstructure noise or jumps

This section shows that with proper adjustments, the estimator $\widehat{\sigma}^2_{n,f}(\bar{t})$ can be employed for the analysis of a more general data generating process where prices are affected by microstructure noise or can display a finite (ω-dependent) number of jumps on $[0, T]$, two important aspects that play a relevant role in the study of financial time series.

3.1 Robustness to microstructure effects

The following results emphasize the suitability of our theoretical framework to deal with microstructure noise effects in the observed data. Zu and Boswijk [56] study an estimator in our class, specifically employing a rectangular kernel. Consider logarithmic asset prices X_{t_i} that are observed at discrete times $t_0, \dots, t_n$ and are subject to an observation error due to microstructure noise. For simplicity, the times t_i are taken equispaced, and the noise is assumed to be additive i.i.d.

Assumption 3.1 Assume that observations are equally spaced ($\varDelta_i = \overline{\varDelta}_n$). Let

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

where Y_{t_i} is the unobservable efficient price satisfying Assumption 2.1, and ε_i denotes the noise component. The noise process $(\varepsilon_i)_{i=0,1,\dots,n}$ is i.i.d. and independent of Y with $\mathrm{E}[\varepsilon_i] = 0$ and $\mathrm{E}[\varepsilon_i^8] < +\infty$.

In what follows, we set by $V_\varepsilon = \mathrm{E}[\varepsilon_i^2]$ and $\kappa_\varepsilon = \mathrm{E}[\varepsilon_i^4]$.

Lemma 3.2 *Let Assumptions 2.4 and 3.1 hold. If* $R_n^{(\sigma^2)}(\bar{t}) = o_p\left(\sqrt{\overline{\varDelta}_n f_n(0)}\right)$ *and* $f_n(0)\overline{\varDelta}_n \to 0$ *as* $n \to \infty$, *then*

$$\frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left(\frac{1}{2}\overline{\Delta}_n\widehat{\sigma}^2_{n,f}(\bar{t}) - V_\varepsilon\right) \longrightarrow \mathbf{N}\left(0, \frac{1}{2}c_f(\kappa_\varepsilon + V_\varepsilon^2)\right), \tag{3.1}$$

where the convergence is in distribution.

It is immediate to see that the market microstructure-induced bias is given by

$$\mathrm{E}\big[\widehat{\sigma}^2_{n,f}(\bar{t}) - \sigma^2(\bar{t})\big] = \frac{2V_\varepsilon}{\overline{\Delta}_n} + o\left(\frac{1}{\overline{\Delta}_n}\right),$$

which diverges at rate n. However, Lemma 3.2 states that when appropriately corrected by a factor $\frac{1}{2}\overline{\Delta}_n$, a consistent estimate of the noise variance is given by

$$\widehat{V}_\varepsilon = \frac{1}{2}\overline{\Delta}_n\widehat{\sigma}^2_{n,f}(\bar{t}).$$

To obtain a consistent and asymptotically mixed normally distributed estimator of the spot variance, we follow the two-scale approach in Zhang et al. [55] and propose an estimator that also uses overlapping prices at a lower frequency. The idea is to remove the market microstructure noise by subtracting volatility estimators constructed with observation records at two different frequencies, leaving the latent volatility unaffected. The approach we are proposing here is simple but not efficient. Efficient estimation could be achieved, in principle, by using multiscales [54], or by smoothing the observed time series via pre-averaging as in Jacod et al. [26], or by using autocovariances and a flat-top kernel as in Barndorff-Nielsen et al. [10]. Define an integer $\overline{n} < n$ and set

$$\widehat{\sigma}^{2,TS}_{n,\overline{n}}(\bar{t}) = \frac{1}{\overline{n}}\sum_{i=1}^{n-\overline{n}+1} f_n(t_{i-1} - \bar{t})\big((X_{t_{i+\overline{n}-1}} - X_{t_{i-1}})^2 - (X_{t_i} - X_{t_{i-1}})^2\big). \tag{3.2}$$

The following theorem shows that $\widehat{\sigma}^{2,TS}_{n,\overline{n}}(\bar{t})$ is a consistent and asymptotically mixed normally distributed estimator in the presence of microstructure noise.

Theorem 3.3 *Let Assumptions 2.4 and 3.1 hold. As $n \to \infty$, let $f_n(0) \to \infty$ and $\overline{n} \to \infty$ in such a way that $\overline{n} f_n(0)\overline{\Delta}_n \to 0$, $R_n^{(\sigma^2)}(\bar{t}) = o_p(1)$, and $L_n\sqrt{\overline{\Delta}_n\overline{n}/f_n(0)} \to 0$. Then we have*

$$\widehat{\sigma}^{2,TS}_{n,\overline{n}}(\bar{t}) \xrightarrow{\mathrm{P}} (\sigma^2)^\star_{\bar{t}}.$$

Furthermore, if $R_n^{(\sigma^2)}(\bar{t}) = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n\overline{n}}\right)$ and $\overline{n} = c(\overline{\Delta}_n)^{-\frac{2}{3}}$ with $c \in \mathbb{R}$, then

$$\frac{1}{\sqrt{f_n(0)(\overline{\Delta}_n)^{\frac{1}{3}}}}\left(\widehat{\sigma}^{2,TS}_{n,\overline{n}}(\bar{t}) - (\sigma^2)^\star_{\bar{t}}\right) \xrightarrow{\mathcal{L}-(s)} \mathbf{MN}\left(0, 2c_f\left(V_\varepsilon^2 + c(\sigma^4)^\star_{\bar{t}}\right)\right),$$

where the convergence is stable in law.

Notice that the loss in estimation efficiency when passing from a framework without noise to a framework with noise is similar to that experienced when estimating integrated volatility with the two-scale estimator, in the presence of noise, instead of realized variance in the absence of noise.

3.2 Robustness to jumps

We now consider the case where a doubly stochastic Poisson jump process is added to the stochastic integral driving the state variable dynamics.

Assumption 3.4 The adapted process (X_t) defined on $[0, T]$ satisfies

$$X_t = Y_t + J_t,$$

where (Y_t) fulfils Assumption 2.1 and $dJ_t = c_J(t)\, dN_t$, with a nonexplosive Poisson counting process (N_t) whose intensity is an adapted stochastic process (λ_t), the sizes of the jumps occur at times $\tau_1, \dots, \tau_{N_t}$, and the jumps sizes are given by i.i.d. random variables $c_J(\tau_j)$ such that $\mathrm{P}[c_J(\tau_j) = 0] = 0$ for $j = 1, \dots, N_t$.

Following the approach in Mancini [37], we define our estimator to be

$$\widehat{\sigma}^2_{n,f}(\bar{t}) = \sum_{i=1}^{n} f_n(t_{i-1} - \bar{t})\, (\Delta X_i)^2\, I_{\{(\Delta X_i)^2 \le \vartheta_n\}}, \tag{3.3}$$

where (ϑ_n) is a suitable sequence. The aim of the threshold ϑ_n is to disentangle the discontinuous variation induced by the Poisson jumps on $(t_{i-1}, t_i]$ from the continuous variation induced by the Brownian motion. Asymptotically, disentangling is possible when (ϑ_n) converges to zero more slowly than the modulus of continuity of the Brownian paths, as specified in the next theorem. Note that (ϑ_n) can also be either a function of time or a stochastic process (see Mancini and Renò [39]). The special case of (3.3) with the indicator kernel is used in Aït-Sahalia and Jacod [1]. An alternative, but not efficient, option is the locally averaged bipower variation proposed by Veraart [50]. Both approaches admit infinite jump activity in the data-generating process.

Theorem 3.5 *Let Assumptions 2.2, 2.4 and 3.4 hold. As $n \to \infty$, let $f_n(0) \to \infty$ and $\vartheta_n \to 0$ in such a way that $f_n(0)\overline{\Delta}_n \to 0$ and $\vartheta_n / (\overline{\Delta}_n \log \frac{1}{\overline{\Delta}_n}) \to \infty$. Then we have*

$$\widehat{\sigma}^2_{n,f}(\bar{t}) \xrightarrow{\mathrm{P}} \sigma^2(\bar{t}).$$

Furthermore, if $R_n^{(\sigma^2)}(\bar{t}) = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n}\right)$, then

$$\frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}} \left(\widehat{\sigma}^2_{n,f}(\bar{t}) - (\sigma^2)^\star_{\bar{t}}\right) \xrightarrow{\mathcal{L}-(s)} \mathbf{MN}\left(0, 2c_f \left(H'\sigma^4\right)^\star_{\bar{t}}\right),$$

where the convergence is stable in law.

4 Relation to the Fourier estimator

In this section, we analyze the Fourier estimator of spot volatility first introduced in Malliavin and Mancino [35]. In particular, we show that the Fourier estimator can be written as the sum of a delta sequence estimator, where the function $f_n(\cdot)$ is set to be equal to the Fejér sequence (see Example 2), and a zero-mean noisy term.

In the Fourier method, classical harmonic analysis is combined with stochastic calculus to connect the Fourier transform of the log-price process (X_t) to the Fourier transform of the volatility function σ_t^2. Specifically, the spot volatility estimator is defined to be

$$\widehat{\sigma}^{2,F}_{n,n',N}(\bar{t}) = \sum_{|k|\leq N} \left(1 - \frac{|k|}{N}\right) \mathcal{H}_{n,n'}(k) e^{\mathrm{i}k\tau}, \tag{4.1}$$

where

$$\mathcal{H}_{n,n'}(k) := \frac{T}{2n'+1} \sum_{|s|\leq n'} \mathcal{F}_n(dX)(s)\mathcal{F}_n(dX)(k-s), \tag{4.2}$$

and

$$\mathcal{F}_n(dX)(s) := \frac{1}{T}\sum_{j=1}^{n} e^{-\mathrm{i}s\tau_{j-1}} \Delta X_j \tag{4.3}$$

is the discrete Fourier transform of dX_t. Here $\tau = 2\pi\bar{t}/T$ and $\tau_i = 2\pi t_i/T$ are rescaled times. Still the observations of X are not necessarily evenly spaced. Malliavin and Mancino [36] prove that when X and σ have continuous paths and $N = n'$,

$$\widehat{\sigma}^{2,F}_{n,n'=N,N}(\bar{t}) \xrightarrow{\mathrm{P}} \sigma^2_{\bar{t}}$$

as $n, N \to \infty$. They also provide a weak convergence result for a Lebesgue average of $\widehat{\sigma}^{2,F}_{n,n'=N,N}(\bar{t}) - \sigma_t^2$ on $[0, T]$, but no central limit theorems for the estimation error of the spot variance are known.

In order to apply the Fourier estimator, it is necessary to choose the number n' of Fourier coefficients of the price process used in the computation of the volatility coefficients, and the number N of volatility coefficients used in the reconstruction of the volatility trajectory. Both n' and N are sequences depending on n. Importantly, here we do not restrict to the choice $n' = N$, suggesting that a higher n' is beneficial. A reference value for equally spaced data is $n' = n/2$, also known as Nyquist frequency (see Priestley [48, Sect. 7]).

In what follows, we show that the Fourier estimator does not belong directly to our class but can be rearranged into the sum of two terms: the volatility estimator $\widehat{\sigma}^2_{n,f}(\bar{t})$, where (f_n) is a rescaled Fejér sequence, and a cross-product term with zero mean.

Proposition 4.1 *The Fourier estimator given in* (4.1) *is such that*

$$\widehat{\sigma}^{2,F}_{n,n',N}(\bar{t}) = \widehat{\sigma}^2_{n,f}(\bar{t}) + \phi_{n,f,g}(\bar{t}),$$

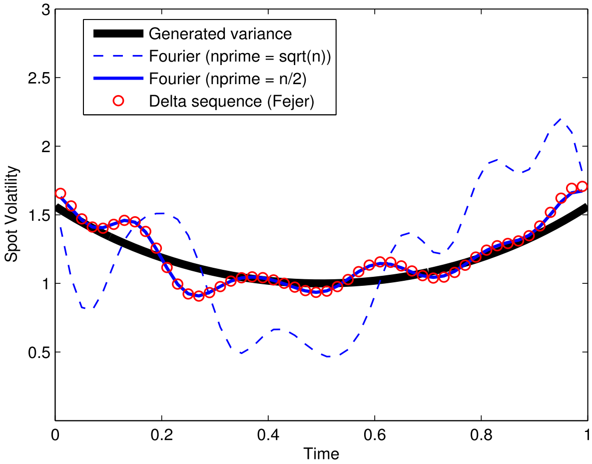

Fig. 2 Fourier estimates of the spot volatility of a single simulated path (the *thick solid line*) with $n = 2500$ and $N = 8$, in the cases $n' = n/2$ (*thin solid line*) and $n' = \sqrt{n}$ (*dashed line*). It is clear that a lower n' leads to a higher variance. We also report the estimator (2.3) with the Fejér delta sequence; it is almost identical to the Fourier estimator with $n' = n/2$ but computationally faster

where

$$\widehat{\sigma}^2_{n,f}(\bar{t}) = \sum_{i=1}^{n} f_n(t_{i-1} - \bar{t})(\Delta X_i)^2$$

and

$$\phi_{n,f,g}(\bar{t}) = \frac{1}{g_n(0)} \sum_{i=1}^{n} \sum_{i \neq j=1}^{n} f_n(t_{j-1} - \bar{t}) g_n(t_{j-1} - t_{i-1}) \Delta X_i \Delta X_j,$$

with $f_n(x) = \frac{1}{T} F_{N-1}(2\pi x/T)$ *and* $g_n(x) = \frac{1}{T} D_{n'}(2\pi x/T)$. *It holds that* $\mathrm{E}[\phi_{n,f,g}(\bar{t})] = 0$, *and if* σ *is independent from* W, *the covariance between* $\widehat{\sigma}^2_{n,f}(\bar{t})$ *and* $\phi_{n,f,g}(\bar{t})$ *is zero.*

Proposition 4.1 is relevant because it tells us that, at least when σ is independent from W, we have $\mathrm{Var}(\widehat{\sigma}^{2,F}_{n,n',N}(\bar{t}) - (\sigma^2)^{\star}_{\bar{t}}) = \mathrm{Var}(\widehat{\sigma}^2_{n,f}(\bar{t}) - (\sigma^2)^{\star}_{\bar{t}}) + \mathrm{Var}(\phi_{n,f,g}(\bar{t}))$. Since $\mathrm{Var}(\phi_{n,f,g}(\bar{t})) \geq 0$, in the absence of noise (e.g., by sampling sparsely), the Fourier estimator is never more efficient than the Fejér one and even requires a higher computational burden. The cross-product term $\phi_{n,f,g}(\bar{t})$ in fact typically adds noise (see Fig. 2). Kanatani [33] made a similar remark in the case of integrated volatility estimation. Note that in contrast, the cross-product term might be beneficial to the reduction of the mean-square error in the presence of market microstructure noise; see Mancino and Sanfelici [40, 41] and Barndorff-Nielsen et al. [10].

The above findings are clearly illustrated in Fig. 2. The spot volatility estimates obtained with the Fourier method are negligibly different from those obtained with the delta sequence (whose computational burden is substantially lower) when $n' = n/2$ and display a larger estimation error when $n' = \sqrt{n}$. Further simulation evidence suggests that in the unequally spaced case, the optimal choice of n' is $\frac{n}{2H'(\bar{t})}$.

Finally, we explicitly state a CLT for the Fejér delta sequence (which does not satisfy property (ii) of Assumption 2.4).

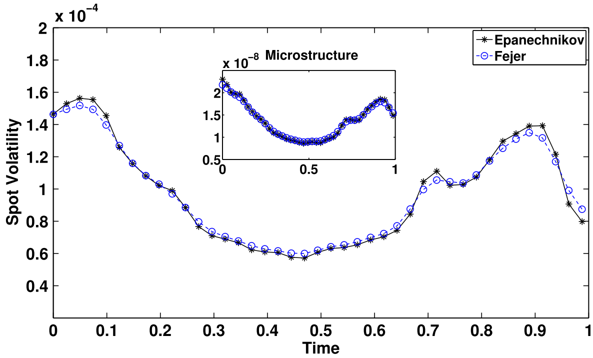

Fig. 3 Intraday spot volatility for the S&P 500 stock index futures over one year of data calculated using the two-scale estimator (3.2) with used observation frequencies of 1 minute and 5 seconds. We use two delta sequences: Epanechnikov kernel and Fejér. Days with relevant jump activity are previously removed from the sample. The *inset* shows the average estimate of the microstructure noise variance V_ε

Proposition 4.2 *Define* $\widehat{\sigma}^2_{n,f}(\bar{t})$ *as in* (2.3) *with* $f_n(x) = \frac{1}{2\pi}F_N(x)$, *where* F_N *is defined in* (2.13), *with* $x \in (-\pi, \pi)$. *Assume that Assumptions* 2.1, 2.2, *and* 2.4 *hold, with the exception of property* (ii) *in Assumption* 2.4, *and also that* $H'(s)$ *is piecewise Liptschitz on* $[0, T]$. *Assume now that as* $n, N \to \infty$, $N/n \to 0$ *in such a way that there exists a sequence* $\varepsilon_n \to 0$ *such that* $N^3\varepsilon_n^4/n \to \infty$ *and* $N/(n\varepsilon_n^\Gamma) \to \infty$. *Then, if* $\bar{t} \in (0, T)$, *then*

$$\frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left(\widehat{\sigma}^2_{n,f}(\bar{t}) - (\sigma^2)^\star_{\bar{t}}\right) \xrightarrow{\mathcal{L}-(s)} \mathbf{MN}\left(0, \frac{4}{3}(H'\sigma^4)^\star_{\bar{t}}\right).$$

Note that in the case $\Gamma = 1$, choosing for the Fejér sequence for example $\varepsilon_n = N_n^{-\alpha}$ and $N_n = n^\delta$ with $0 < \alpha < 1/2$ and $\delta \in (1/2, 1)$ sufficiently close to 1 fulfils the requirements in the statement of Proposition 4.2 for (n, N_n, ε_n), which also imply the requirements in Proposition 2.5.

5 Empirical application

In this final section, we apply the proposed estimators (3.2) and (3.3) to a set of market data consisting of high-frequency transactions of the S&P 500 stock index futures. We restrict our attention to the year 1999 and to contracts closer to maturity. Transactions are recorded over 251 trading days between 8.30 a.m. to 3.15 p.m. and interpolated to a 5-second grid. Every day, we then have a total of 4,860 price returns. We use two delta sequences: a kernel one (Epanechnikov), with $h = 15$ minutes, and a trigonometric one (Fejér, (2.13)), with $N = 81$.

To calculate the low frequency part on the right-hand side of equality (3.2), we apply a subsampling technique similar to that described in Zhang et al. [55] with $\overline{n} = 12$, which corresponds to one-minute returns. In order to avoid the effect of jump dynamics in the observed data, we first remove from the sample all the days characterized by significant price changes using the procedure described below. Figure 3 plots the estimated intraday spot volatility averaged across days and calculated in daily time units. The difference between the estimates obtained with the two delta sequences is small. The well known U-shape is clearly detected, as already observed

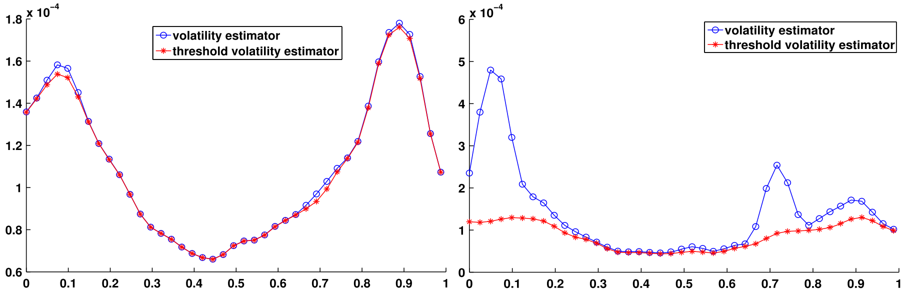

Fig. 4 Estimated intraday spot volatility for the S&P 500 stock index futures averaged over one year of data calculated using the original volatility estimator (2.3) and the threshold estimator (3.3). The delta sequence is that obtained with the Fejér delta sequence with $N = 81$. *Left panel*: jump-filtered 5-minute series. *Right panel*: sample made of the 28 days characterized by large price movements. The significance level of jump detection is set to 99 %. The volatility is measured in daily units

in previous studies; see, for instance, Andersen and Bollerslev [3]. The estimate of the microstructure variance V_ε obtained with both delta sequences is also provided.

We now turn to the jump-robust estimator (3.3), and we use 5-minute returns for computation of spot volatility estimators, to soften the impact of microstructure noise. To show that our threshold estimator $\widehat{\sigma}^2_{n,f}(\bar{t})$ is robust to price jumps, we compare it with the original spot volatility estimator (2.3) using the 5-minute data set created by removing all days with relevant jump activity. The resulting intraday volatility curves then should be almost identical. To identify the jumps, we employ the $C - Tz$ statistics in Corsi et al. [13]. After setting the daily significance level of a jump to 99 %, a total of 28 days are detected and then excluded from the sample. The left panel in Fig. 4 shows that the volatility curves obtained with estimators (2.3) applied to the filtered series and (3.3) to the complete series match almost everywhere, meaning that $\widehat{\sigma}^2_{n,f}(\bar{t})$ is not affected by large price movements and is able to provide robust estimates of the intraday volatility dynamics. We then apply both (2.3) and (3.3) to a sample made of the 28 days initially removed; the result is plotted in Fig. 4, right panel. As expected, now the two curves behave quite differently, especially around the market opening time. The reported figures have been obtained with the Fejér delta sequence, but very similar results have been obtained with the Epanechnikov kernel.

6 Conclusions

We study a wide and unifying class of spot volatility estimators constructed using delta sequences. The class includes kernel estimators as a special case, as well as estimators constructed with nonkernel delta sequences (e.g., the Fejér one). Under mild hypotheses on the data-generating process, we provide a full asymptotic theory for the estimators within the class, and we propose suitable modifications to allow for the presence of microstructure noise or price discontinuities. Our contribution makes different estimation techniques comparable in terms of efficiency. For example, we relate the Fourier estimator of spot volatility with the estimator obtained with the Fejér delta sequence, showing that the latter is more convenient, at least in the case in which the price follows a Brownian semimartingale and there is neither leverage effect nor microstructure noise. We finally apply specific delta sequences (Epanechnikov and Fejér) to a data set of high-frequency stock index futures and successfully recover the traditional U-shaped intraday volatility pattern.

It is well known that in a nonparametric setting, the choice of the efficient estimator for spot volatility is model-dependent. Having enlarged the class of spot volatility estimators to include various approaches used in the literature (e.g., kernels, trigonometric delta functions, and wavelets) could be helpful for future research, in solving the problem of finding an efficient delta sequence given the assumed mathematical framework. Indeed, our generalization on the one hand gives more chances to find an efficient delta sequence for a given model and on the other hand provides a tool for their comparison.

The paper further leaves open the possibility of additional theoretical developments, such as studying the joint contribution of microstructure noise and jumps, possibly using the techniques in Jacod et al. [26] and Podolskij and Vetter [46, 47], and finding the asymptotic distribution of the Fourier estimator of spot volatility. We leave all these interesting issues for future research.

On the empirical side, spot volatility estimation is growing in importance in applied work in many fields, such as model selection for asset price dynamics (e.g., when testing for jumps) and option pricing. Further, there are relevant open issues in financial econometrics, such as specifying the dynamics of intraday effects, in which precise estimation of spot volatility is crucial. We also leave more empirical work for further research.

Acknowledgements We are grateful to Simona Sanfelici, Claudio Pacati, Jean Jacod, Kris Boudt, Fulvio Corsi, Maria Elvira Mancino, Imma Curato, and Davide Pirino for valuable comments and helpful discussions. We benefited from funds given by the GNAMPA group of INdAM (Istituto Nazionale di Alta Matematica) would like to dedicate this work to the memory of Prof. Paul Malliavin.

Appendix A: Proofs

In what follows, we use $\int(\dots)\,dx$ to denote an integral over $\mathbb{R}$. Moreover, C or K indicates a constant which does not depend on i, nor on the sequence $F=\{f_n, n\in\mathbb{N}\}$, but can depend on $\bar{t}$ and the localizing sequence (τ_m), and which keeps the same name even when changing from line to line or from one side to another of an inequality. Without loss of generality, we assume that $\sigma\geq 0$. Recall that for any Lebesgue-integrable function a and for all $\ell\geq 1$, we have, by Jensen's inequality,

$$\frac{1}{\Delta_i}\int_{t_{i-1}}^{t_i}|a_s|\,ds \leq \left(\frac{1}{\Delta_i}\int_{t_{i-1}}^{t_i}|a_s|^{\ell}\,ds\right)^{\frac{1}{\ell}}. \tag{A.1}$$

By a localization procedure (as in [25], Sect. 3.6.3), in all the following proofs we can assume that μ and σ are bounded (as (ω,t) vary within $\Omega\times[0,T]$).

We use the notations O_p and o_p that should be understood as follows. For two (possibly random) sequences (A_n) and (B_n), we say that $A_n=O_p(B_n)$ if there exists $\bar{n}$ such that for all $n\geq\bar{n}$, we have that for any $\varepsilon>0$, there exists a constant $\eta>0$

such that $\mathrm{P}[|A_n| > \eta |B_n|] < \varepsilon$. When the $|B_n|$ are a.s. positive, $A_n = O_p(B_n)$ means that for sufficiently large n, (A_n/B_n) is a sequence bounded in probability. With (b_n) a sequence of positive reals tending to zero as $n \to \infty$, we say that $A_n = o_p(b_n)$ if $A_n/b_n \to 0$ in probability, and $A_n = O_{a.s.}(b_n)$ if for almost all fixed ω, we have $A_n = O(b_n)$, that is, there exists $\eta(\omega) > 0$ such that for sufficiently large n, we have $|A_n| \le \eta b_n$.

In order to prove Theorem 2.7, we use Lemma A.1 below several times with A being equal to σ^k for some powers $k \in \{0, 1, 2, 3, 4\}$. Note that by property (2.1) and the boundedness of σ, we have, for $k \ge 2$,

$$\mathrm{E}_{u\wedge s}\big[|\sigma_u - \sigma_s|^k\big] = \mathrm{E}_{u\wedge s}\big[|\sigma_u - \sigma_s|^2 |\sigma_u - \sigma_s|^{k-2}\big] \le C|u-s|^{\Gamma}, \tag{A.2}$$

and similarly, we also have, for $k \ge 2$,

$$\mathrm{E}_{u\wedge s}\big[|\sigma_u^2 - \sigma_s^2|^k\big] \le C|u-s|^{\Gamma}. \tag{A.3}$$

For $k = 1$, we have instead by Jensen's inequality $\mathrm{E}_{u\wedge s}[|\sigma_u - \sigma_s|] \le C|u-s|^{\Gamma/2}$ and $\mathrm{E}_{u\wedge s}[|\sigma_u^2 - \sigma_s^2|] \le C|u-s|^{\Gamma/2}$.

Proof of Proposition 2.5 Assume first that $0 < \bar{t} < T$. Using the boundedness of σ^2 and properties (2.8) and (2.9), we can write

$$\begin{aligned} R_n^{(\sigma^2)}(\bar{t}) &= \int_0^T f_n(s-\bar{t})\sigma_s^2\,ds - \left(\int_0^T f_n(s-\bar{t})\,ds + o(1)\right)(\sigma^2)_{\bar{t}}^{\star} \\ &= \int_0^T f_n(s-\bar{t})\big(\sigma_s^2 - (\sigma^2)_{\bar{t}}^{\star}\big)\,ds + o_p(1) \end{aligned}$$

and, using (2.1) and property (2.9),

$$\begin{aligned} \mathrm{E}\left[\left|\int_0^T f_n(s-\bar{t})\big(\sigma_s^2 - (\sigma^2)_{\bar{t}}^{\star}\big)\,ds\right|\right] &\le \int_0^T f_n(s-\bar{t})\mathrm{E}\big[\big|\sigma_s^2 - (\sigma^2)_{\bar{t}}^{\star}\big|\big]\,ds \\ &= \int_{|s-\bar{t}|<\varepsilon_n} f_n(s-\bar{t})\mathrm{E}\big[\big|\sigma_s^2 - (\sigma^2)_{\bar{t}}^{\star}\big|\big]\,ds \\ &\quad + \int_{|s-\bar{t}|\ge\varepsilon_n} f_n(s-\bar{t})\mathrm{E}\big[\big|\sigma_s^2 - (\sigma^2)_{\bar{t}}^{\star}\big|\big]\,ds \\ &\le C\varepsilon_n^{\Gamma/2} + C\int_{|x|\ge\varepsilon_n} f_n(x)\,dx \longrightarrow 0. \end{aligned}$$

If instead $\bar{t} = T$, we repeat the same reasoning as before using that $\int_{T-\varepsilon<s<T} f_n(s-T)ds \to \psi_f^-$ and we proceed in a similar way if $\bar{t} = 0$. This proves (2.4). To prove (2.5) for f_n, it is thus enough to prove that $g_n(x) = \frac{f_n^2(x)}{c_f f_n(0)}$ satisfies (2.8), which is straightforward from property (2.10), and (2.9), which is obtained, using $\sup_x f_n(x) \le C f_n(0)$, since

$$0 \le \int_{|x|\ge\varepsilon_n} g_n(x)\,dx = \frac{1}{c_f}\int_{|x|\ge\varepsilon_n} f_n(x)\frac{f_n(x)}{f_n(0)}\,dx \le C\int_{|x|\ge\varepsilon_n} f_n(x)\,dx \longrightarrow 0.$$

To prove (2.6), use the boundedness of σ^2 and (2.10) and write

$$\int_0^T \frac{f_n^4(t-\bar{t})}{f_n^2(0)}\sigma_s^2 dt \le C\int_0^T \frac{f_n^2(t-\bar{t})}{f_n(0)} f_n(0)\, dt \le C f_n(0). \qquad \square$$

Lemma A.1 (i) *For a sequence of processes $(A^{(n)})$ bounded by the same constant K, if f_n are Lipschitz functions and* $\max_i \Delta_i = O(\overline{\Delta}_n)$, *then*

$$\int_0^T f_n(s-\bar{t})A_s^{(n)}\, ds - \sum_{i=1}^n f_n(t_{i-1}-\bar{t})\int_{t_{i-1}}^{t_i} A_s^{(n)}\, ds = O_{a.s.}(L_n\overline{\Delta}_n). \tag{A.4}$$

As a corollary, under (2.4) *and if $L_n\overline{\Delta}_n \to 0$, we have $\sum_{i=1}^n f_n(t_{i-1}-\bar{t})\Delta_i \to 1$ as $n\to\infty$.*

(ii) *Consider a bounded càdlàg process A satisfying* (2.1). *Assume that* $\sup_{x\in D}|f_n(x)| \le C f_n(0)$. *If either $f_n \ge 0$ or $\overline{\Delta}_n^{\Gamma/2}\int_0^T |f_n(s-\bar{t})|\, ds \to 0$ and if* (2.7) *holds, if both f_n and $g_n = f_n^2/(c_f f_n(0))$ satisfy* (2.4), *under* $\max_i \Delta_i = O(\overline{\Delta}_n)$ *and* (2.2), *we have*

$$\frac{1}{\overline{\Delta}_n}\sum_{i=1}^n f_n(t_{i-1}-\bar{t})A(t_{i-1})\Delta_i^2 \xrightarrow[n\to\infty]{\mathrm{P}} (H'A)^\star_{\bar{t}}, \tag{A.5}$$

$$\frac{1}{\overline{\Delta}_n f_n(0)}\sum_{i=1}^n f_n^2(t_{i-1}-\bar{t})A(t_{i-1})\Delta_i^2 \xrightarrow[n\to\infty]{\mathrm{P}} c_f(H'A)^\star_{\bar{t}}. \tag{A.6}$$

(iii) *Under* (2.1) *and the boundedness of σ, for any $p = 0, 1, 2, 3$, there exists $\alpha > 0$ such that for all $i = 1,\dots,n$ and for a suitable constant C_p,*

$$\mathrm{E}_{i-1}\left[|\sigma_{t_{i-1}}\Delta W_i|^p \left|\int_{t_{i-1}}^{t_i}\left(\sigma_s - \sigma_{t_{i-1}}\right) dW_s\right|^{4-p}\right] \le C_p\Delta_i^2\sigma_{t_{i-1}}^p\left(\Delta_i^\alpha I_{\{t_{i-1}\in B_\varepsilon(\bar{t})\}} + I_{\{t_{i-1}\notin B_\varepsilon(\bar{t})\}}\right).$$

(iv) *Under the same assumptions as for* (ii) *above, for any bounded càdlàg process M and $\alpha > 0$, we have*

$$\sum_{i=1}^n \frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)\overline{\Delta}_n} M_{t_{i-1}}\Delta_i^2\left(\Delta_i^\alpha I_{\{t_{i-1}\in B_\varepsilon(\bar{t})\}} + I_{\{t_{i-1}\notin B_\varepsilon(\bar{t})\}}\right) \xrightarrow{\mathrm{P}} 0.$$

(v) *Under* (2.1) *and the boundedness of σ, for any $p = 0, 1$, we have for all $i = 1,\dots,n$ and for some $\alpha > 0$,*

$$\mathrm{E}_{i-1}\left[|\sigma_{t_{i-1}}\Delta W_i|^p \left|\int_{t_{i-1}}^{t_i}\left(\sigma_s - \sigma_{t_{i-1}}\right) dW_s\right|^{2-p}\right] \le C_p\Delta_i\sigma_{t_{i-1}}^p\left(\Delta_i^\alpha I_{\{t_{i-1}\in B_\varepsilon(\bar{t})\}} + I_{\{t_{i-1}\notin B_\varepsilon(\bar{t})\}}\right).$$

(vi) *Under* (2.1) *and the boundedness of* σ, *for* $p \in [1, 8]$,

$$\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} |\sigma_s^2 - \sigma_{t_{i-1}}^2| \, ds\right)^p\right] \le K \Delta_i^p \big(\Delta_i^{\Gamma/2} I_{\{t_{i-1} \in B_\varepsilon(\bar{t})\}} + I_{\{t_{i-1} \notin B_\varepsilon(\bar{t})\}}\big).$$

Proof See the online available Appendix B. □

Proof of Theorem 2.7 It is not restrictive to set $\mu_t \equiv 0$, as will be shown at the end of the proof. We start by proving the stated convergence in law. Using (2.4) and then (A.4) and that $\mu \equiv 0$, we have

$$\begin{aligned}
&\frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left(\hat{\sigma}_{n,f}^2(\bar{t}) - (\sigma^2)_{\bar{t}}^{\star}\right) \\
&= \frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left(\sum_{i=1}^n f_n(t_{i-1} - \bar{t})\Delta(X_i)^2 - (\sigma^2)_{\bar{t}}^{\star}\right) \\
&= \frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left(\sum_{i=1}^n f_n(t_{i-1} - \bar{t})\Delta(X_i)^2 - \int_0^T f_n(s - \bar{t})\sigma^2(s)\, ds - R_n^{(\sigma^2)}(\bar{t})\right) \\
&= \frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left(\sum_{i=1}^n f_n(t_{i-1} - \bar{t})\left(\Delta(X_i)^2 - \int_{t_{i-1}}^{t_i} \sigma^2(s)\, ds\right)\right. \\
&\quad \left. + O_{a.s.}(L_n\overline{\Delta}_n) - R_n^{(\sigma^2)}(\bar{t})\right) \\
&= \sum_{i=1}^n U_i + O_{a.s.}\left(L_n\sqrt{\frac{\overline{\Delta}_n}{f_n(0)}}\right) - \frac{R_n^{(\sigma^2)}(\bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}},
\end{aligned} \tag{A.7}$$

where for $i = 1, \ldots, n$,

$$U_i := \frac{f_n(t_{i-1} - \bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}}\left(\left(\int_{t_{i-1}}^{t_i} \sigma_s\, dW_s\right)^2 - \int_{t_{i-1}}^{t_i} \sigma_s^2\, ds\right). \tag{A.8}$$

Since we assumed that $L_n\sqrt{\frac{\overline{\Delta}_n}{f_n(0)}} \to 0$ and $R_n^{(\sigma^2)}(\bar{t}) = o_p(\sqrt{f_n(0)\overline{\Delta}_n})$, the last two terms in (A.7) tend to zero in probability, and thus it is sufficient to derive a central limit theorem stable in law for $\sum_{i=1}^n U_i$. For that, we refer to Theorem IX.7.28 in Jacod and Shiryaev [29], ensuring that the following are sufficient conditions:

$$\text{(i)} \quad \sum_{i=1}^n \mathrm{E}_{i-1}[U_i] \xrightarrow{\mathrm{P}} 0, \qquad \text{(ii)} \quad \sum_{i=1}^n \mathrm{E}_{i-1}[U_i^2] \xrightarrow{\mathrm{P}} V_{\bar{t}},$$

$$\text{(iii)} \quad \sum_{i=1}^n \mathrm{E}_{i-1}[U_i^4] \xrightarrow{\mathrm{P}} 0, \qquad \text{(iv)} \quad \textstyle\sum_{i=1}^n \mathrm{E}_{i-1}[U_i \Delta Z_i] \xrightarrow{\mathrm{P}} 0,$$

where $\mathrm{E}_{i-1}[\cdot]$ abbreviates $\mathrm{E}[\cdot|\mathcal{F}_{t_{i-1}}]$, and (iv) has to hold in both cases where $Z = W$ or $Z = B$, with B any bounded martingale orthogonal (in the martingales sense) to W. Condition (i) is immediately proved, using the Itô isometry, via

$$\sum_{i=1}^{n} \mathrm{E}_{i-1}[U_i] = \sum_{i=1}^{n} \frac{f_n(t_{i-1}-\bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}} \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s\, dW_s\right)^2 - \int_{t_{i-1}}^{t_i} \sigma_s^2\, ds\right] = 0.$$

As for condition (ii), consider

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{E}_{i-1}[U_i^2] &= \sum_{i=1}^{n} \frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)\overline{\Delta}_n} \mathrm{E}_{i-1}\left[\left(\left(\int_{t_{i-1}}^{t_i} \sigma_s\, dW_s\right)^2 - \int_{t_{i-1}}^{t_i} \sigma_s^2\, ds\right)^2\right] \\ &= \sum_{i=1}^{n} \frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)\overline{\Delta}_n} \left(\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s\, dW_s\right)^4\right] + \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s^2\, ds\right)^2\right]\right. \\ &\quad \left. - 2\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s\, dW_s\right)^2 \left(\int_{t_{i-1}}^{t_i} \sigma_s^2\, ds\right)\right]\right). \end{aligned} \tag{A.9}$$

All three conditional expectations contain some leading terms, which we need to compute exactly. Basically, for $s \in (t_{i-1}, t_i]$, we write $\sigma_s = \sigma_{t_{i-1}} + (\sigma_s - \sigma_{t_{i-1}})$, we find exact equalities for the expressions containing $\sigma_{t_{i-1}}$, and by using (2.1) we show that the other terms are asymptotically negligible. Now write

$$\begin{aligned}&\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s\, dW_s\right)^4\right] \\ &= \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} (\sigma_{t_{i-1}} + \sigma_s - \sigma_{t_{i-1}})\, dW_s\right)^4\right] \\ &= 3\sigma_{t_{i-1}}^4 \Delta_i^2 + \sum_{p=0}^{3} c_p \mathrm{E}_{i-1}\left[(\sigma_{t_{i-1}} \Delta W_i)^p \left(\int_{t_{i-1}}^{t_i} (\sigma_s - \sigma_{t_{i-1}})\, dW_s\right)^{4-p}\right],\end{aligned}$$

with suitable constants c_p. Using Lemma A.1(iii) and (iv), the first term within brackets in (A.9) contributes by

$$\sum_{i=1}^{n} \frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)\overline{\Delta}_n} 3\sigma_{t_{i-1}}^4 \Delta_i^2 + o_p(1),$$

and thanks to (A.6), this in turn converges in probability to $3c_f (H'\sigma^4)^{\star}_{\bar{t}}$.

As for the second term within brackets in (A.9), we similarly decompose it as

$$\begin{aligned}&\mathrm{E}_{i-1}\left[\left(\sigma_{t_{i-1}}^2 \Delta_i + \int_{t_{i-1}}^{t_i} \sigma_s^2 - \sigma_{t_{i-1}}^2\, ds\right)^2\right] \\ &\quad = \sigma_{t_{i-1}}^4 \Delta_i^2 + \sum_{q=0}^{1} c_q \mathrm{E}_{i-1}\left[(\sigma_{t_{i-1}}^2 \Delta_i)^q \left(\int_{t_{i-1}}^{t_i} (\sigma_s^2 - \sigma_{t_{i-1}}^2)\, ds\right)^{2-q}\right].\end{aligned}$$

Using Lemma A.1(vi) with $p=2-q$, both terms with $q=0,1$ are bounded by

$$\sigma^{2q}_{t_{i-1}}\Delta_i^q\Delta_i^{2-q}\big(\Delta_i^\alpha I_{\{t_{i-1}\in B_\varepsilon(\bar t)\}}+I_{\{t_{i-1}\notin B_\varepsilon(\bar t)\}}\big)$$

for some $\alpha>0$, which by Lemma A.1(iv) give asymptotically negligible contributions to (A.9), so that the second term within brackets in (A.9) contributes by

$$\sum_{i=1}^n\frac{f_n^2(t_{i-1}-\bar t)}{f_n(0)\overline{\Delta}_n}\sigma^4_{t_{i-1}}\Delta_i^2+o_p(1)\longrightarrow c_f(H'\sigma^4)^\star_{\bar t}$$

by (A.6).

As for the third term within brackets in (A.9), we still decompose it as

$$\begin{aligned}
&\mathrm{E}_{i-1}\bigg[\bigg(\sigma_{t_{i-1}}\Delta W_i+\int_{t_{i-1}}^{t_i}(\sigma_s-\sigma_{t_{i-1}})\,dW_s\bigg)^2\bigg(\sigma^2_{t_{i-1}}\Delta_i+\int_{t_{i-1}}^{t_i}(\sigma_s^2-\sigma^2_{t_{i-1}})\,ds\bigg)\bigg]\\
&\quad=\sigma^4_{t_{i-1}}\Delta_i^2+\sum_{p=0}^1c_p\sigma^{2+p}_{t_{i-1}}\Delta_i\mathrm{E}_{i-1}\bigg[(\Delta W_i)^p\bigg(\int_{t_{i-1}}^{t_i}(\sigma_s-\sigma_{t_{i-1}})\,dW_s\bigg)^{2-p}\bigg]\\
&\qquad+\sigma^2_{t_{i-1}}\mathrm{E}_{i-1}\bigg[(\Delta W_i)^2\int_{t_{i-1}}^{t_i}(\sigma_s^2-\sigma^2_{t_{i-1}})\,ds\bigg]\\
&\qquad+\sum_{q=0}^1c_q\sigma^q_{t_{i-1}}\mathrm{E}_{i-1}\bigg[(\Delta W_i)^q\bigg(\int_{t_{i-1}}^{t_i}(\sigma_s-\sigma_{t_{i-1}})\,dW_s\bigg)^{2-q}\int_{t_{i-1}}^{t_i}(\sigma_s^2-\sigma^2_{t_{i-1}})\,ds\bigg].
\end{aligned}$$

By Lemma A.1(v) and (iv), the terms with $p=0,1$ give asymptotically negligible contributions to (A.9). Noting that $\int_{t_{i-1}}^{t_i}|\sigma_s^2-\sigma^2_{t_{i-1}}|\,ds\le K\Delta_i$, the terms with $q=0,1$ are reduced to terms of exactly the same type as the ones with $p=0,1$ before and thus are asymptotically negligible. Now we deal with the term

$$\sigma^2_{t_{i-1}}\mathrm{E}_{i-1}\bigg[(\Delta W_i)^2\int_{t_{i-1}}^{t_i}(\sigma_s^2-\sigma^2_{t_{i-1}})\,ds\bigg],$$

which by the Hölder inequality is dominated by

$$\sigma^2_{t_{i-1}}\sqrt{\mathrm{E}_{i-1}[(\Delta W_i)^4]}\sqrt{\mathrm{E}_{i-1}\bigg[\bigg|\int_{t_{i-1}}^{t_i}(\sigma_s^2-\sigma^2_{t_{i-1}})\,ds\bigg|^2\bigg]}.$$

Using Lemma A.1(vi) with $p=2$, we obtain in turn that this is less than

$$K\Delta_i^2\big(\Delta_i^\alpha I_{\{t_{i-1}\in B_\varepsilon(\bar t)\}}+I_{\{t_{i-1}\notin B_\varepsilon(\bar t)\}}\big)$$

with a suitable $\alpha>0$, and by Lemma A.1(iv) also the contribution of this term is asymptotically negligible. Therefore, the third term within brackets in (A.9) has the same limit in probability as

$$-2\sum_{i=1}^n\frac{f_n^2(t_{i-1}-\bar t)}{f_n(0)\overline{\Delta}_n}\sigma^4_{t_{i-1}}\Delta_i^2\xrightarrow{\mathrm{P}}-2c_f(H'\sigma^4)^\star_{\bar t}\tag{A.10}$$

by Lemma A.1(ii). Summing up, the probability limit in condition (ii) is $V_{\bar{t}} = 2c_f(H'\sigma^4)^{\star}_{\bar{t}}$.

We now deal with condition (iii), where we only have to check the negligibility of the fourth conditional moments, and even some rough estimates are sufficient. So we write

$$\sum_{i=1}^n \mathrm{E}_{i-1}[U_i^4] = \sum_{i=1}^n \frac{f_n^4(t_{i-1}-\bar{t})}{f_n^2(0)\overline{\Delta}_n^2} \mathrm{E}_{i-1}\Big[\Big(\Big(\int_{t_{i-1}}^{t_i} \sigma_s\,dW_s\Big)^2 - \int_{t_{i-1}}^{t_i} \sigma_s^2\,ds\Big)^4\Big]$$
$$\leq \sum_{i=1}^n \frac{f_n^4(t_{i-1}-\bar{t})}{f_n^2(0)\overline{\Delta}_n^2} \mathrm{E}_{i-1}\Big[\Big(\int_{t_{i-1}}^{t_i} \sigma_s\,dW_s\Big)^8 + \Big(\int_{t_{i-1}}^{t_i} \sigma_s^2\,ds\Big)^4\Big].$$

As σ is assumed to be bounded, we have by the BDG inequality that $\mathrm{E}_{i-1}[(\int_{t_{i-1}}^{t_i} \sigma_s\,dW_s)^8] \leq K\Delta_i^4$, and the last sum is dominated by

$$K\sum_{i=1}^n \frac{f_n^4(t_{i-1}-\bar{t})\Delta_i}{f_n^2(0)} \frac{\Delta_i^3}{\overline{\Delta}_n^2} \leq K\overline{\Delta}_n \sum_{i=1}^n \frac{f_n^4(t_{i-1}-\bar{t})\Delta_i}{f_n^2(0)}, \tag{A.11}$$

having used that by Assumption 2.2, $\max_i \Delta_i \leq K\overline{\Delta}_n$. However,

$$\overline{\Delta}_n \left| \sum_{i=1}^n \frac{f_n^4(t_{i-1}-\bar{t})\Delta_i}{f_n^2(0)} - \frac{\int_0^T f_n^4(s-\bar{t})\,ds}{f_n^2(0)} \right| \xrightarrow{\mathrm{P}} 0.$$

In fact, $f_n^4 = g(f_n)$ with $g(x) = x^4$ and by assumption $|f_n| \leq Cf_n(0)$; thus, f_n^4 is Lipschitz in a neighborhood of 0 (say $B_\varepsilon(0)$) where f_n is Lipschitz, and it has Lipschitz constant $Kf_n(0)^3L_n$. So the sum in the last display is bounded by

$$\frac{\overline{\Delta}_n}{f_n^2(0)} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \big|f_n^4(t_{i-1}-\bar{t}) - f_n^4(s-\bar{t})\big|\,ds\big(I_{t_i\in B_\varepsilon(\bar{t})} + I_{t_i\in B_\varepsilon^c(\bar{t})}\big).$$

For sufficiently small $\max_i \Delta_i$, if $t_i \in B_\varepsilon(\bar{t})$, then also $t_{i-1} \in B_\varepsilon(\bar{t})$. For $s \in (t_{i-1}, t_i]$, we thus have $t_{i-1}-\bar{t}, s-\bar{t} \in B_\varepsilon(0)$, and

$$\frac{\overline{\Delta}_n}{f_n^2(0)} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \big|f_n^4(t_{i-1}-\bar{t}) - f_n^4(s-\bar{t})\big|\,ds\,I_{t_i\in B_\varepsilon(\bar{t})} \leq K\frac{\overline{\Delta}_n^2}{f_n^2(0)} f_n^3(0)L_n$$
$$\leq Kf_n(0)\overline{\Delta}_n\overline{\Delta}_nL_n \longrightarrow 0$$

by the assumptions on L_n, $\overline{\Delta}_n$, and $f_n(0)$. On the other hand, if $t_i \in B_\varepsilon^c(\bar{t})$ and n is sufficiently large, then for $s \in (t_{i-1}, t_i]$, we have $t_{i-1}-\bar{t}, s-\bar{t} \in B_\varepsilon^c(0)$, and by assumption $|f_n(t_{i-1}-\bar{t})|, |f_n(s-\bar{t})| \leq M_\varepsilon$; thus,

$$\frac{\overline{\Delta}_n}{f_n^2(0)} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \big|f_n^4(t_{i-1}-\bar{t}) - f_n^4(s-\bar{t})\big|\,ds\,I_{t_i\in B_\varepsilon^c(\bar{t})} \leq K\frac{\overline{\Delta}_n}{f_n(0)} M_\varepsilon^4 \longrightarrow 0$$

since $f_n(0) \to \infty$ by assumption. Consequently, using (2.6) with $\sigma \equiv 1$, the limit in probability of (A.11) is the same as for

$$K\bar{\Delta}\frac{\int_0^T f_n^4(s-\bar{t})ds}{f_n^2(0)} \leq K\overline{\Delta}_n f_n(0) \longrightarrow 0.$$

We finally consider condition (iv), starting from the case $Z = B$. Denote $R_t = \int_0^t \sigma_s \, dW_s$, $C_t = \int_0^t \sigma_s^2 \, ds$, and $M = R^2 - C$. Orthogonality (in the martingales sense) of two martingales $M^{(1)}$ and $M^{(2)}$ means that $[M^{(1)}, M^{(2)}] \equiv 0$. The fact that $[B, W] \equiv 0$ implies that B is orthogonal also to R and M since $d(R_t^2) = 2R_t \, dR_t + \sigma_t^2 \, dt$. However, $[M, B] \equiv 0$ means that MB is a martingale, which implies that $\mathrm{E}_{i-1}[M_{t_i} B_{t_i}] = M_{t_{i-1}} B_{t_{i-1}}$. Using this and the fact that B and M are martingales, we have that $\mathrm{E}_{t_{i-1}}[\Delta M_i \Delta B_i] = 0$ for $i = 1, \dots, n$, and hence

$$\mathrm{E}_{t_{i-1}}[U_i \Delta B_i] = \frac{f_n(t_{i-1} - \bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}} \mathrm{E}_{t_{i-1}}[\Delta M_i \Delta B_i] = 0.$$

Therefore, condition (iv) is verified.

When instead $Z = W$, we have

$$\begin{aligned}
&\sum_{i=1}^n \mathrm{E}_{i-1}\left[U_i \Delta W_i\right] \\
&\quad \leq \sum_{i=1}^n \frac{f_n(t_{i-1} - \bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}} \sqrt{\mathrm{E}_{i-1}\left[\left(\left(\int_{t_{i-1}}^{t_i} \sigma_s \, dW_s\right)^2 - \int_{t_{i-1}}^{t_i} \sigma_s^2 \, ds\right)^2\right]} \sqrt{\mathrm{E}_{i-1}[(\Delta W_i^2)]}.
\end{aligned}$$

Reasoning as in (A.9) and below, we have

$$\begin{aligned}
&\mathrm{E}_{i-1}\left[\left(\left(\int_{t_{i-1}}^{t_i} \sigma_s \, dW_s\right)^2 - \int_{t_{i-1}}^{t_i} \sigma_s^2 \, ds\right)^2\right] \\
&\quad \leq 2\sigma_{t_{i-1}}^4 \Delta_i^2 + K\Delta_i^2 \left(\Delta_i^\alpha I_{\{t_{i-1} \in B_\varepsilon(\bar{t})\}} + I_{\{t_{i-1} \notin B_\varepsilon(\bar{t})\}}\right)
\end{aligned}$$

for a suitable $\alpha > 0$. Therefore, the last term within the above display is dominated by

$$\begin{aligned}
&\sum_{i=1}^n \frac{f_n(t_{i-1} - \bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}} \Delta_i^{\frac{3}{2}} \sqrt{2\sigma_{t_{i-1}}^4 + K\left(\Delta_i^\alpha I_{\{t_{i-1} \in B_\varepsilon(\bar{t})\}} + I_{\{t_{i-1} \notin B_\varepsilon(\bar{t})\}}\right)} \\
&\quad \leq K \sum_{i=1}^n \frac{f_n(t_{i-1} - \bar{t})}{\sqrt{f_n(0)}} \Delta_i.
\end{aligned}$$

By (A.4) and the assumptions $L_n \overline{\Delta}_n \to 0$ and $f_n(0) \to \infty$, this last has the same probability limit as $\frac{\int_0^T f_n(s-\bar{t})\,ds}{\sqrt{f_n(0)}}$, which is zero as $\int_0^T f_n(s - \bar{t})\,ds \to 1$ by (2.4), and condition (iv) is verified also when $Z = W$. This completes the proof of the stable convergence of $(\widehat{\sigma}^2_{n,F}(\bar{t}))$.

For the convergence in probability, the condition $R_n^{(\sigma^2)}(\bar{t}) = o_p(\sqrt{f_n(0)\overline{\Delta}_n})$ is not required. Indeed, by multiplying both sides of (A.7) by $\sqrt{f_n(0)\overline{\Delta}_n}$, we find

$$\begin{aligned}\hat{\sigma}^2_{n,f}(\bar{t}) - (\sigma^2)^\star_{\bar{t}} = &\sum_{i=1}^n f_n(t_{i-1} - \bar{t})\left(\left(\int_{t_{i-1}}^{t_i} \sigma_s\,dW_s\right)^2 - \int_{t_{i-1}}^{t_i} \sigma_s^2\,ds\right)\\ &+ O_{a.s.}\big(f_n(0)\overline{\Delta}_n\big) - R_n^{(\sigma^2)}(\bar{t}).\end{aligned}$$

The last two terms are $o_p(1)$ by the assumption $f_n(0)\overline{\Delta}_n \to 0$ and (2.4), whereas we check the negligibility of the first term by using the law of large numbers for the sum of martingale differences (see, e.g., Lemma 4.1 in Jacod [25]). It is sufficient to show

$$\sum_i \mathrm{E}_{i-1}\left[\left(\sqrt{f_n(0)\overline{\Delta}_n}U_i\right)^2\right] = f_n(0)\overline{\Delta}_n \sum_i \mathrm{E}_{i-1}[U_i^2] \longrightarrow 0,$$

which is ensured by $\sum_i \mathrm{E}_{i-1}[U_i^2] \to 2c_f(H'\sigma^4)^\star_{\bar{t}}$ obtained with the computations for (ii) and by $f_n(0)\overline{\Delta}_n \to 0$.

We now show that when $\mu \not\equiv 0$, the previous results of consistency and asymptotic mixed normality of $\hat{\sigma}^2(\bar{t}) - \sigma^2(\bar{t})$ still hold since the contribution of $\int \mu_s\,ds$ to the estimation error $\hat{\sigma}^2_{n,F}(\bar{t}) - \sigma^2(\bar{t})$ is negligible with respect to the Brownian term. If $\mu \not\equiv 0$, by following (A.7) and substituting ΔX_i we have

$$\begin{aligned}\frac{\hat{\sigma}^2_{n,f}(\bar{t}) - (\sigma^2)^\star_{\bar{t}}}{\sqrt{f_n(0)\overline{\Delta}_n}} = &\sum_{i=1}^n U_i + 2\sum_i \frac{f_n(t_{i-1} - \bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}} \int_{t_{i-1}}^{t_i} \sigma_s\,dW_s \int_{t_{i-1}}^{t_i} \mu_s\,ds\\ &+ \sum_i \frac{f_n(t_{i-1} - \bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}}\left(\int_{t_{i-1}}^{t_i} \mu_s\,ds\right)^2\\ &+ O_{a.s.}\left(L_n\sqrt{\frac{\overline{\Delta}_n}{f_n(0)}}\right) + \frac{R_n^{(\sigma^2)}(\bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}},\end{aligned}$$

and we see firstly that again the assumption $\frac{R_n^{(\sigma^2)}(\bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}} \xrightarrow{\mathrm{P}} 0$ is only needed for the CLT and secondly that showing the negligibility of the second and third terms is sufficient also to state the consistency of $\hat{\sigma}^2_{n,f}(\bar{t})$. To deal with those terms, which are both of the type $\sum_i \xi_i$, we apply Lemma 4.1 in [25], and in both cases we check condition (4.4) from [25]. Since $\sum_i \mathrm{E}_{i-1}[|\xi_i|] \le \sum_i \sqrt{\mathrm{E}_{i-1}[|\xi_i|^2]} \le K\sqrt{\sum_i \mathrm{E}_{i-1}[|\xi_i|^2]}$, it is sufficient to check that $\sum_i \mathrm{E}_{i-1}[\xi_i^2] \to 0$. By the boundedness of μ and the BDG inequality and then (A.6), we have

$$\sum_i \frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)\overline{\Delta}_n} \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma_s\, dW_s\right)^2 \left(\int_{t_{i-1}}^{t_i} \mu_s\, ds\right)^2\right]$$
$$\leq K \sum_i \frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)\overline{\Delta}_n} \Delta_i^3$$
$$\leq K\overline{\Delta}_n \sum_i \frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)\overline{\Delta}_n} \Delta_i^2 \xrightarrow{\mathrm{P}} 0$$

and

$$\sum_i \frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)\overline{\Delta}_n} \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \mu_s\, ds\right)^4\right] \leq K\overline{\Delta}_n^2 \sum_i \frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)\overline{\Delta}_n} \Delta_i^2 \xrightarrow{\mathrm{P}} 0,$$

as desired. □

Proof of Remark 2.8 Assume that $\bar{t} \in (0, T)$. For the Gaussian kernel, with ε as in Assumption 2.1, we have

$$\frac{R_n^{(\sigma^2)}(\bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}} = A_n + \frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}} \left(\frac{1}{h_n} \int_{s\in[0,T]:|s-\bar{t}|\leq\varepsilon} K\left(\frac{s-\bar{t}}{h_n}\right) \sigma_s^2\, ds - (\sigma^2)_{\bar{t}}^{\star}\right) \tag{A.12}$$

with $A_n = \frac{1}{h_n\sqrt{f_n(0)\overline{\Delta}_n}} \int_{s\in[0,T]:|s-\bar{t}|>\varepsilon} K(\frac{s-\bar{t}}{h})\sigma_s^2\, ds$. By changing variables via $x = (s-\bar{t})/h_n$, the second term in (A.12) is written as $B_n + C_n + D_n + E_n$ with

$$B_n = \frac{\int_{-\varepsilon/h_n}^{0} K(x)(\sigma^2_{\bar{t}-h_n|x|} - \sigma^2_{\bar{t}-})\, dx}{\sqrt{f_n(0)\overline{\Delta}_n}}, \qquad C_n = \frac{\int_0^{\varepsilon/h_n} K(x)(\sigma^2_{\bar{t}+h_n x} - \sigma^2_{\bar{t}})\, dx}{\sqrt{f_n(0)\overline{\Delta}_n}},$$
$$D_n = -\frac{\int_{-\infty}^{-\varepsilon/h_n} K(x)\sigma^2_{\bar{t}-}\, dx}{\sqrt{f_n(0)\overline{\Delta}_n}}, \qquad \bar{E}_n = -\frac{\int_{\varepsilon/h_n}^{+\infty} K(x)\sigma^2_{\bar{t}}\, dx}{\sqrt{f_n(0)\overline{\Delta}_n}}.$$

Using (2.1), we have

$$\mathrm{E}[|B_n|] \leq \frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}} \int_{-\varepsilon/h_n}^{0} K(x)|xh_n|^{\Gamma/2}\, dx \leq C\sqrt{nh_n^{1+\Gamma}} \longrightarrow 0.$$

A similar result holds for C_n. Moreover,

$$
\begin{aligned}
|A_n| &= \frac{1}{h_n\sqrt{f_n(0)\overline{\Delta}_n}}\left|\int_{s\in[0,T]:|s-\bar{t}|>\varepsilon} K\left(\frac{s-\bar{t}}{h}\right)\sigma_s^2\,ds\right| \\
&= \sqrt{\frac{h_n}{K(0)\overline{\Delta}_n}}\left|\int_{-\bar{t}/h_n}^{-\varepsilon/h_n} K(x)\sigma^2_{h_n x+\bar{t}}\,dx + \int_{\varepsilon/h_n}^{(T-\bar{t})/h_n} K(x)\sigma^2_{h_n x+\bar{t}}\,dx\right| \\
&\le C\left[\sqrt{nh_n}\int_{-\bar{t}/h_n}^{-\varepsilon/h_n} K(x)\,dx + \sqrt{nh_n}\int_{\varepsilon/h_n}^{(T-\bar{t})/h_n} K(x)\,dx\right].
\end{aligned}
$$

For the first term, on $[-\bar{t}/h_n, -\varepsilon/h_n]$ for small h_n, we have the estimate $K(x) \le K(-\varepsilon/h_n) = Ce^{-\frac{1}{2}(\frac{\varepsilon}{h_n})^2}$, and since ε is fixed, this last quantity is $o(h_n^4)$ so that $\sqrt{nh_n}\int_{-\bar{t}/h_n}^{-\varepsilon/h_n} K(x)\,dx \le C\sqrt{nh_n^7} \to 0$. The second term can be dealt with in a similar way.

Finally for D_n, the well-known inequality

$$
\int_y^{+\infty} K(x)\,dx \le \frac{C}{y}K(y) \quad \text{for } y > 0
$$

and the boundedness of σ imply

$$
\begin{aligned}
|D_n| &= \frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\int_{-\infty}^{-\varepsilon/h_n} K(x)\sigma^2_{\bar{t}-}\,dx \\
&\le C\sqrt{nh_n}\int_{-\infty}^{-\varepsilon/h_n} K(x)\,dx \le C\sqrt{nh_n}\frac{h_n}{\varepsilon}e^{\frac{-\varepsilon^2}{2h_n^2}} \\
&\le C\sqrt{nh_n^{11}} \longrightarrow 0.
\end{aligned}
$$

The term E_n is dealt with similarly.

For the Epanechnikov kernel, $A_n = 0$ for h small enough, and the rest is similar.

For the indicator kernel, for n large enough,

$$
\mathrm{E}\left[\frac{|R_n^{(\sigma^2)}(\bar{t})|}{\sqrt{f_n(0)\overline{\Delta}_n}}\right] \le \frac{\int_{|x|\le 1/f_n(0)} \frac{1}{2}f_n(0)\mathrm{E}[|\sigma^2_{\bar{t}+x/2} - (\sigma^2)^\star_{\bar{t}}|]\,dx}{\sqrt{f_n(0)\overline{\Delta}_n}} \le C\frac{f_n(0)^{-\Gamma/2}}{\sqrt{f_n(0)\overline{\Delta}_n}},
$$

providing the same result. □

Proof of Lemma 3.2 See the online available Appendix B. □

Proof of Theorem 3.3 See the online available Appendix B. □

Proof of Theorem 3.5 Write $X = Y + J$, where Y is a continuous semimartingale. By Theorem 1 in Mancini [37], for n large enough, we can write almost surely

$$\widehat{\sigma}^2_{n,f}(\bar{t}) = \sum_{i=1}^{n} f_n(t_{i-1} - \bar{t})(\Delta Y_i)^2 - \sum_{i=1}^{n} f_n(t_{i-1} - \bar{t})(\Delta Y_i)^2 I_{\{\Delta N_i \neq 0\}}.$$

Theorem 2.7 can be applied to the first term, whereas the second term is $O_p(N_T \overline{\Delta}_n f_n(0))$ then it is $o_p\left(\sqrt{\overline{\Delta}_n f_n(0)}\right)$, where N is the Poisson counting process. So the second term is vanishing in the limit. □

Lemma A.2 *The result* (A.6) *continues to hold when* (f_n) *is given by the Fejér sequence*, *the observations of* X *are not necessarily evenly spaced*, H' *is piecewise Liptshitz on* $[0, T]$ *and* $H'A$ *satisfies* (2.1).

Proof First, remark that for any bounded process A, we have

$$\begin{aligned}
&\sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left| f_n(s - \bar{t}) - f_n(t_{j-1} - \bar{t}) \right| |A_s| \, ds \\
&\quad \leq C \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left| \sum_{|k| \leq N} \left(1 - \frac{|k|}{N+1}\right) \left(e^{\mathrm{i}k(s-\bar{t})} - e^{\mathrm{i}k(t_{j-1}-\bar{t})}\right) \right| ds \\
&\quad \leq C \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left| \sum_{|k| \leq N} \left(1 - \frac{|k|}{N+1}\right) e^{\mathrm{i}k(s-\bar{t})} \mathrm{i}k\Delta_j \right| ds \\
&\quad \leq C\overline{\Delta}_n \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} f_n(s - \bar{t}) N \, ds \\
&\quad \leq C N \overline{\Delta}_n \to 0. \qquad\qquad (A.13)
\end{aligned}$$

It follows from $\sup_x |f_n(x)| \leq C f_n(0)$ that

$$\begin{aligned}
&\frac{1}{f_n(0)} \left| \sum_{j=1}^{n} f_n^2(t_{j-1} - \bar{t}) \int_{t_{j-1}}^{t_j} A_s \, ds - \int_0^T f_n^2(s - \bar{t}) A_s \, ds \right| \\
&= \frac{1}{f_n(0)} \left| \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left(f_n^2(t_{j-1} - \bar{t}) - f_n^2(s - \bar{t})\right) A_s \, ds \right| \qquad\qquad (A.14) \\
&= \frac{1}{f_n(0)} \left| \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left(f_n(t_{j-1} - \bar{t}) - f_n(s - \bar{t})\right)\left(f_n(t_{j-1} - \bar{t}) + f_n(s - \bar{t})\right) A_s \, ds \right| \\
&\leq C \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left| f_n(t_{j-1} - \bar{t}) - f_n(s - \bar{t}) \right| |A_s| \, ds \\
&\leq C N \overline{\Delta}_n. \qquad\qquad (A.15)
\end{aligned}$$

Therefore, for the Fejér sequence, in place of (A.4) we make use of

$$\frac{1}{f_n(0)}\left|\sum_{j=1}^{n} f_n^2(t_{j-1}-\bar{t})\int_{t_{j-1}}^{t_j} A_s\,ds-\int_0^T f_n^2(s-\bar{t})A_s\,ds\right|\leq CN\overline{\Delta}_n. \tag{A.16}$$

Secondly, as commented after (2.13), f_n is a delta sequence, and thus (2.5) holds. In order to show (A.6), we check that the following tends to 0 in probability:

$$\begin{aligned}
&\frac{1}{f_n(0)}\sum_{i=1}^{n} f_n^2(t_{i-1}-\bar{t})A_{t_{i-1}}\frac{\Delta_i^2}{\overline{\Delta}_n}-c_f(H'A)_{\bar{t}}^{\star}\\
&=\frac{1}{f_n(0)}\sum_{i=1}^{n} f_n^2(t_{i-1}-\bar{t})A_{t_{i-1}}\frac{\Delta_i^2}{\overline{\Delta}_n}-\frac{1}{f_n(0)}\int_0^T f_n^2(s-\bar{t})(H'A)_s\,ds+o_P(1)\\
&=-A_n-B_n+o_P(1),
\end{aligned}$$

where

$$\begin{aligned}
A_n&=\frac{1}{f_n(0)}\int_0^T\sum_i\big(f_n^2(s-\bar{t})-f_n^2(t_{i-1}-\bar{t})\big)I_{(t_{i-1},t_i]}(s)(H'A)_s\,ds,\\
B_n&=\frac{1}{f_n(0)}\int_0^T\sum_i f_n^2(t_{i-1}-\bar{t})\big((H'A)_s-A_{t_{i-1}}\Delta_i/\overline{\Delta}_n\big)I_{(t_{i-1},t_i]}(s)\,ds.
\end{aligned}$$

Now A_n coincides with (A.14), where $H'A$ here plays the role of A there, and tends to zero a.s. As for B_n, split it as $B_n=B_{1,n}+B_{2,n}$, where

$$\begin{aligned}
B_{1,n}&=\frac{1}{f_n(0)}\sum_i\int_{t_{i-1}}^{t_i} f_n^2(t_{i-1}-\bar{t})H_s'(A_s-A_{t_{i-1}})\,ds,\\
B_{2,n}&=\frac{1}{f_n(0)}\sum_i\int_{t_{i-1}}^{t_i} f_n^2(t_{i-1}-\bar{t})(H_s'-\Delta_i/\overline{\Delta}_n)A_{t_{i-1}}\,ds.
\end{aligned}$$

For $B_{2,n}$, use that $\frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)}\leq C|f_n(t_{i-1}-\bar{t})|=f_n(t_{i-1}-\bar{t})$, condition (2.2) and that H' is piecewise Lipschitz, which imply that, for $i=1,\dots,n$, $\sup_{s\in]t_{i-1},t_i]}|H'(s)-\Delta_i/\overline{\Delta}_n|\leq K\overline{\Delta}_n$, and the boundedness of A, and note that from (A.13) we have

$$\sum_i f_n(t_{i-1}-\bar{t})\Delta_i=\int_0^T f_n(s-\bar{t})\,ds+O_P(N\overline{\Delta}_n)\to 1.$$

So we get

$$|B_{2,n}|\leq\sum_i\int_{t_{i-1}}^{t_i} f_n(t_{i-1}-\bar{t})C\bar{\Delta}_n\,ds\leq C\overline{\Delta}_n\sum_i f_n(t_{i-1}-\bar{t})\Delta_i\longrightarrow 0.$$

For $B_{1,n}$, we proceed analogously as from (B.2) in the online available Appendix B, to the end of the proof of part (ii) of Lemma A.1 by replacing $f_n(t_{i-1}-\bar{t})$ in (B.2) with $\frac{f_n^2(t_{i-1}-\bar{t})}{f_n(0)}$, and then replacing (A.4) with (A.16). □

Proof of Proposition 4.1 From (4.1)–(4.3) we have

$$\begin{aligned}
\widehat{\sigma}^{2,F}_{n,n',N}(\bar{t}) &= \sum_{|k|\le N}\left(1-\frac{|k|}{N}\right)\left(\frac{T}{2n'+1}\sum_{|s|\le n'}\mathcal{F}_n(dX)(s)\mathcal{F}_n(dX)(k-s)\right)e^{ik\tau}\\
&= \frac{1}{T}\sum_{|k|\le N}\left(1-\frac{|k|}{N}\right)\\
&\quad\times\left(\frac{1}{2n'+1}\sum_{|s|\le n'}\sum_{j'=1}^{n}\sum_{j=1}^{n}e^{-\mathrm{i}s\tau_{j'-1}}e^{-\mathrm{i}(k-s)\tau_{j-1}}\Delta X_j\Delta X_{j'}\right)e^{\mathrm{i}k\tau}\\
&= \frac{1}{T}\sum_{|k|\le N}\left(1-\frac{|k|}{N}\right)\left[\frac{1}{2n'+1}\sum_{|s|\le n'}\left(\sum_{j=1}^{n}e^{-\mathrm{i}k\tau_{j-1}}(\Delta X_j)^2\right.\right.\\
&\quad\left.\left.+\sum_{j'\ne j=1}^{n}e^{-\mathrm{i}s\tau_{j'-1}}e^{-\mathrm{i}(k-s)\tau_{j-1}}\Delta X_{j'}\Delta X_j\right)\right]e^{\mathrm{i}k\tau}\\
&= \frac{1}{T}\sum_{j=1}^{n}\sum_{|k|\le N}\left(1-\frac{|k|}{N}\right)e^{\mathrm{i}k(\tau-\tau_{j-1})}(\Delta X_j)^2\\
&\quad+\frac{1}{T(2n'+1)}\sum_{j'\ne j=1}^{n}\sum_{|k|\le N}\left(1-\frac{|k|}{N}\right)e^{\mathrm{i}k(\tau-\tau_{j-1})}\\
&\quad\times\sum_{|s|\le n'}e^{\mathrm{i}s(\tau_{j-1}-\tau_{j'-1})}\Delta X_{j'}\Delta X_j.
\end{aligned}$$

By the definitions of the Dirichlet and Fejér sequences in (2.12) and (2.13) we then get the first statement of the proposition. The second statement is straightforward. □

Proof of Proposition 4.2 As in the proof of Theorem 2.7, we can assume that $\mu=0$. Write

$$\sigma^2_{n,f}(\bar{t})-(\sigma^2)^\star_{\bar{t}}=TE+DE+ME,$$

where

$$\begin{aligned}
TE &= \text{truncation error} = \int_0^T f_n(s-\bar{t})\sigma_s^2\,ds-(\sigma^2)^\star_{\bar{t}},\\
DE &= \text{discretization error} = \sum_{i=1}^{n}f_n(t_{i-1}-\bar{t})\int_{t_{i-1}}^{t_i}\sigma_s^2\,ds-\int_0^T f_n(s-\bar{t})\sigma_s^2\,ds,\\
ME &= \text{martingale error} = \sum_{i=1}^{n}f_n(t_{i-1}-\bar{t})(\Delta X_i)^2-\sum_{i=1}^{n}f_n(t_{i-1}-\bar{t})\int_{t_{i-1}}^{t_i}\sigma_s^2\,ds.
\end{aligned}$$

For the martingale error, we do the same as in the proof of Theorem 2.7 to show that

$$\frac{ME}{\sqrt{\bar{\Delta}_n f_n(0)}} \to \mathbf{MN}\Big(0, \frac{4}{3}(\sigma^4 H')^{\star}_{\bar{t}}\Big),$$

using also Lemma A.2. For the term DE, (A.16) gives, for a suitable constant C,

$$|DE| \leq CN\overline{\Delta}_n;$$

thus, $\frac{1}{\sqrt{\bar{\Delta}_n f_n(0)}} DE = \sqrt{\frac{n}{N}} DE \leq C\sqrt{\frac{N}{n}} \to 0$ a.s., and DE is negligible.

For the term TE, with no loss of generality we can take $T = \pi$, so that we have $\bar{t} \in (0, \pi)$. For large n, we then have

$$\begin{aligned}
TE = R_n^{(\sigma^2)} &= \int_0^T f_n(s-\bar{t})\sigma_s^2\,ds - (\sigma^2)^{\star}_{\bar{t}} \\
&= \int_{-\bar{t}}^{\pi-\bar{t}} f_n(x)\sigma^2_{x+\bar{t}}\,dx - \int_{-\pi}^{\pi} f_n(x)(\sigma^2)^{\star}_{\bar{t}}\,dx \\
&= TE1 + TE2 - TE3,
\end{aligned}$$

where

$$\begin{aligned}
TE1 &= \int_{|x|\leq\varepsilon_n} f_n(x)\big(\sigma^2_{x+\bar{t}} - (\sigma^2)^{\star}_{\bar{t}}\big)\,dx, \\
TE2 &= \int_{\{|x|>\varepsilon_n\}\cap(-\bar{t},\pi-\bar{t})} f_n(x)\big(\sigma^2_{x+\bar{t}} - (\sigma^2)^{\star}_{\bar{t}}\big)\,dx, \\
TE3 &= \int_{(-\pi,-\bar{t})\cup(\pi-\bar{t},\pi)} f_n(x)(\sigma^2)^{\star}_{\bar{t}}\,dx.
\end{aligned}$$

For large n, $TE2$ and $TE3$ are both at most $\int_{|x|>\varepsilon_n} f_n(x)\,dx$ by the boundedness of σ, and it turns out that both $TE2$ and $TE3$ are $O_p(1/(N\varepsilon_n^2))$, so that $\frac{1}{\sqrt{\bar{\Delta}_n f_n(0)}}(TE_2 - TE_3) = \sqrt{\frac{n}{N}}(TE2 - TE3) = O_p(\sqrt{n/(N^3\varepsilon_n^4)}) \to 0$. As for $TE1$, we take n such that $\varepsilon_n < \varepsilon$, with ε as in Assumption 2.1, so that for $|x| \leq \varepsilon_n$, we have $\mathrm{E}[|\sigma^2_{x+\bar{t}} - (\sigma^2)^{\star}_{\bar{t}}|] \leq C\varepsilon_n^{\Gamma/2}$, and thus

$$\begin{aligned}
\frac{1}{\sqrt{\bar{\Delta}_n f_n(0)}} E[|TE_1|] = \sqrt{\frac{n}{N}}\mathrm{E}[|TE1|] &\leq \sqrt{\frac{n}{N}} \int_{|x|\leq\varepsilon_n} f_n(x)\mathrm{E}\big[\big|\sigma^2_{x+\bar{t}} - (\sigma^2)^{\star}_{\bar{t}}\big|\big]\,dx \\
&\leq \sqrt{\frac{n\varepsilon_n^{\Gamma}}{N}} \longrightarrow 0,
\end{aligned}$$

which shows that the whole TE is negligible. $\square$

References

1. Aït-Sahalia, Y., Jacod, J.: Testing for jumps in a discretely observed process. Ann. Stat. **37**, 184–222 (2009)
2. Alvarez, A., Panloup, F., Pontier, M., Savy, N.: Estimation of the instantaneous volatility. Stat. Inference Stoch. Process. **15**, 27–59 (2012)
3. Andersen, T., Bollerslev, T.: Intraday periodicity and volatility persistence in financial markets. J. Empir. Finance **4**, 115–158 (1997)
4. Andreou, E., Ghysels, E.: The impact of sampling frequency and volatility estimators on change-point tests. J. Financ. Econom. **2**, 290–318 (2004)
5. Bandi, F., Nguyen, T.: On the functional estimation of jump-diffusion models. J. Econom. **116**, 293–328 (2003)
6. Bandi, F., Phillips, P.: Fully nonparametric estimation of scalar diffusion models. Econometrica **71**, 241–283 (2003)
7. Bandi, F., Renò, R.: Nonparametric stochastic volatility. Working paper (2009). Available at SSRN. doi:10.2139/ssrn.1158438
8. Bandi, F., Renò, R.: Price and volatility co-jumps. J. Financ. Econ. (2015, forthcoming). Available at SSRN. doi:10.2139/ssrn.2014777
9. Bandi, F., Russell, J.: Volatility. In: Linetsky, V., Birge, J. (eds.) Handbooks in Operations Research and Management Science: Financial Engineering, pp. 183–222. Elsevier, Amsterdam (2007), Chap. 5
10. Barndorff-Nielsen, O.E., Hansen, P., Lunde, A., Shephard, N.: Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica **76**, 1481–1536 (2008)
11. Barndorff-Nielsen, O.E., Shephard, N.: Variation, jumps, market frictions and high frequency data in financial econometrics. In: Blundell, R., Newey, W., Persson, T. (eds.) Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress, pp. 328–372. Cambridge University Press, Amsterdam (2007)
12. Boudt, K., Croux, C., Laurent, S.: Outlyingness weighted quadratic covariation. J. Financ. Econom. **9**, 657–684 (2011)
13. Corsi, F., Pirino, D., Renò, R.: Threshold bipower variation and the impact of jumps on volatility forecasting. J. Econom. **159**, 276–288 (2010)
14. Dobrev, D., Andersen, T., Schaumburg, E.: Duration-based volatility estimation. Working paper (2008). Available at http://ideas.repec.org/p/hst/ghsdps/gd08-034.html
15. van Eeden, C.: Mean integrated squared error of kernel estimators when the density and its derivative are not necessarily continuous. Ann. Inst. Stat. Math. **37**, 461–472 (1985)
16. Fan, J., Wang, Y.: Spot volatility estimation for high-frequency data. Stat. Interface **1**, 279–288 (2008)
17. Fan, J., Yao, Q.: Nonlinear Time Series. Springer, Berlin (2003)
18. Florens-Zmirou, D.: On estimating the diffusion coefficient from discrete observations. J. Appl. Probab. **30**, 790–804 (1993)
19. Foster, D., Nelson, D.: Continuous record asymptotics for rolling sample variance estimators. Econometrica **64**, 139–174 (1996)
20. Genon-Catalot, V., Laredo, C., Picard, D.: Non-parametric estimation of the diffusion coefficient by wavelet methods. Scand. J. Stat. **19**, 317–335 (1992)
21. Hoffmann, M., Munk, A., Schmidt-Hieber, J.: Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. Ann. Inst. Henri Poincaré **48**, 1186–1216 (2012)
22. Hoffmann, M.: L_p estimation of the diffusion coefficient. Bernoulli **5**, 447–481 (1999)
23. Hoffmann, M.: Rate of convergence for parametric estimation in a stochastic volatility model. Stoch. Process. Appl. **97**, 147–170 (2002)
24. Jacod, J.: On continuous conditional Gaussian martingales and stable convergence in law. In: Séminaire Proba. XXXI. Lecture Notes in Math., vol. 1655, pp. 232–246. Springer, Berlin (1997)
25. Jacod, J.: Statistics and high-frequency data. In: Kessler, M., Lindner, A., Sørensen, M. (eds.) Statistical Methods for Stochastic Differential Equations, pp. 191–309. Chapman & Hall/CRC Press/Taylor and Francis, London/Boca Raton/London (2012)
26. Jacod, J., Li, Y., Mykland, P., Podolskij, M., Vetter, M.: Microstructure noise in the continuous case: the pre-averaging approach. Stoch. Process. Appl. **119**, 2249–2276 (2009)
27. Jacod, J., Protter, P.: Discretization of Processes. Springer, Berlin (2012)
28. Jacod, J., Rosenbaum, M.: Quarticity and other functionals of volatility: efficient estimation. Ann. Stat. **41**, 1462–1484 (2013)
29. Jacod, J., Shiryaev, A.N.: Limit Theorems for Stochastic Processes. Springer, Berlin (2003)

30. Jacod, J., Todorov, V.: Testing for common arrivals of jumps for discretely observed multidimensional processes. Ann. Stat. **37**, 1792–1838 (2009)
31. Jacod, J., Todorov, V.: Do price and volatility jump together? Ann. Appl. Probab. **20**, 1425–1469 (2010)
32. Johannes, M.: The statistical and economic role of jumps in continuous-time interest rate models. J. Finance **59**, 227–260 (2004)
33. Kanatani, T.: Integrated volatility measuring from unevenly sampled observations. Econ. Bull. **3**, 1–8 (2004)
34. Kristensen, D.: Nonparametric filtering of the realised spot volatility: a kernel-based approach. Econom. Theory **26**, 60–93 (2010)
35. Malliavin, P., Mancino, M.: Fourier series method for measurement of multivariate volatilities. Finance Stoch. **6**, 49–61 (2002)
36. Malliavin, P., Mancino, M.: A Fourier transform method for nonparametric estimation of volatility. Ann. Stat. **37**, 1983–2010 (2009)
37. Mancini, C.: Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scand. J. Stat. **36**, 270–296 (2009)
38. Mancini, C., Calvori, F.J.: Jumps. In: Bauwens, L., Hafner, C., Laurent, S. (eds.) Handbook of Volatility Models and Their Applications, pp. 427–475. Wiley, New York (2012), Chap. 17
39. Mancini, C., Renò, R.: Threshold estimation of Markov models with jumps and interest rate modeling. J. Econom. **160**, 77–92 (2011)
40. Mancino, M., Sanfelici, S.: Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. Comput. Stat. Data Anal. **52**, 2966–2989 (2008)
41. Mancino, M., Sanfelici, S.: Estimating covariance via Fourier method in the presence of asynchronous trading and microstructure noise. J. Financ. Econom. **9**, 367–408 (2011)
42. Mykland, P., Zhang, L.: ANOVA for diffusions and Itô processes. Ann. Stat. **34**, 1931–1963 (2006)
43. Mykland, P., Zhang, L.: Inference for volatility-type objects and implications for hedging. Stat. Interface **1**, 255–278 (2008)
44. Ngo, H.L., Ogawa, S.: A central limit theorem for the functional estimation of the spot volatility. Monte Carlo Methods Appl. **14**, 353–380 (2009)
45. Ogawa, S., Sanfelici, S.: An improved two-step regularization scheme for spot volatility estimation. Econ. Notes **40**, 107–134 (2011)
46. Podolskij, M., Vetter, M.: Bipower-type estimation in a noisy diffusion setting. Stoch. Process. Appl. **119**, 2803–2831 (2009)
47. Podolskij, M., Vetter, M.: Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli **15**, 634–668 (2009)
48. Priestley, M.: Spectral Time Series Analysis. Wiley, Berlin (1979)
49. Renò, R.: Nonparametric estimation of the diffusion coefficient of stochastic volatility models. Econom. Theory **24**, 1174–1206 (2008)
50. Veraart, A.: Inference for the jump part of quadratic variation of Itô semimartingales. Econom. Theory **26**, 331–368 (2010)
51. Walter, G.: Approximation of delta function by wavelets. J. Approx. Theory **71**, 329–343 (1992)
52. Walter, G., Blum, J.: Probability density estimation using delta sequences. Ann. Stat. **6**, 328–340 (1979)
53. Watson, G., Leadbetter, M.: Hazard analysis II. Sankhya, Ser. A **26**, 101–116 (1964)
54. Zhang, L.: Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli **12**, 1019–1043 (2006)
55. Zhang, L., Mykland, P.A., Aït-Sahalia, Y.: A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Am. Stat. Assoc. **100**, 1394–1411 (2005)
56. Zu, Y., Boswijk, P.: Estimating spot volatility with high-frequency financial data. J. Econom. **181**(2), 117–135 (2014)