Decisions in Economics and Finance
https://doi.org/10.1007/s10203-019-00253-y

Estimation of volatility in a high-frequency setting: a short review

Jean Jacod[1]

Received: 12 December 2018 / Accepted: 9 May 2019

Abstract
Our aim is to give an overview of the topic of estimation of volatility, in a high-frequency setting. We emphasize the various possible situations, relative to the underlying process (continuous, or with jumps having finite, or infinite, activity) and to the observation scheme (with microstructure noise or not, under a regular sampling scheme or not). We try to explain a variety of methods, including the most recent ones. Each method is quickly sketched, with comments on its range of applicability. Most results are given in the form of a theorem, with a precise description of the assumptions needed, but of course without proof, and some results are simply mentioned in a somewhat loose way. We consider only the one-dimensional case, although occasional comments are made about the multivariate case. We totally omit the nowadays hot topic when the number of assets is very large, meaning that this number increases as the frequency increases: this is unfortunately not compatible with a "short" review as this one.

Keywords Volatility · High-frequency · Microstructure noise · Fourier methods

JEL Classification C14 · C58 · C60

1 Introduction

We consider the problem of estimating the so-called volatility of a stochastic process indexed by the time $t \geq 0$ and satisfying

$$dX_t = b_t\,dt + \sigma_t\,dW_t + dX_t^J, \tag{1}$$

where W is a standard Brownian motion and X^J is a "pure jump" process, to be described more precisely later. The process X could *a priori* be multidimen-

✉ Jean Jacod
jean.jacod@gmail.com

[1] Institut de Mathématiques de Jussieu, CNRS – UMR 7586, Sorbonne Université, 4 Place Jussieu, 75252 Paris-Cedex 05, France

Published online: 16 May 2019

sional (hence W as well, so σ_t is matrix-valued), but for simplicity we consider the one-dimensional case only, since most features of this case easily extend to the multidimensional case, up to much more cumbersome notation.

The volatility is the process σ_t, or its square $c_t = (\sigma_t)^2$: the terminology is somewhat imprecise about this, and actually in a model like (1) replacing σ_t by its absolute value $|\sigma_t|$ does not change the law of the process, so indeed σ_t and c_t carry the same information, and below we focus on c_t. To be more precise, c_t is called the *spot volatility*, and the *integrated volatility* is

$$C_t = \int_0^t c_s \, \mathrm{d}s.$$

The process X is observed at discrete times over a finite time interval $[0, T]$ and the overall aim is to estimate c_t or C_t (when $t \in [0, T]$ only, of course). As is often the case in statistics, the quality of an estimator is asserted through its asymptotic behavior as the number of observations increases. Put otherwise, we are considering the so-called *high-frequency* setting where, whereas T stays fixed, the number of observations at stage n goes to ∞ as $n \to \infty$ and the maximal distance between successive observations goes to 0.

About spot versus integrated volatility, if $(c_t)_{t\in[0,T]}$ is known, the same is true of $(C_t)_{t\in[0,T]}$, whereas the converse is "almost true": namely, the process C determines the values of c_t except on a Lebesgue-null set of times, and modifying c_t on such a set does not change the process X. Therefore, we mainly focus on the estimation of C_t, and when we want to estimate the spot volatility, we need some additional hypothesis, such as the paths of c being continuous from the right or the left.

We denote as $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space on which X is defined, and the main features of the problem are as follows:

(a) A single path $t \mapsto X_t(\omega)$ of the process X is (partially) observed, hence the estimation target, C_t for example, is really $C_t(\omega)$ for the specific sample point ω giving rise to the observations. We thus have a non-standard statistical problem since the target is itself a random quantity.

(b) If the path $t \mapsto X_t$ were fully observed on $[0, T]$, then C_t would be known for all $t \in [0, T]$, except on a $\mathbb{P}$-null set, because C is actually the continuous part of the quadratic variation of X, which itself is a function of the path of X, again up to a $\mathbb{P}$-null set. This explains why we can hope for estimators which are at least weakly consistent as $n \to \infty$. In contrast, the drift term b_t in (1), or its integrated version $B_t = \int_0^t b_s \, ds$, is not known even when X is fully observed, so there are no consistent estimators (with a fixed horizon T) for them, even when $b_t(\omega) = b$ is a constant.

(c) Quite often, the value X_t at some observation time t is not exactly observed, what we observe is $X_t + \varepsilon_t$, where ε_t is a noise: if the process describes some physical or biological phenomenon, this is the usual observation noise, typically i.i.d. as t varies and most often small; in finance this is the so-called microstructure noise, with a more complicated structure.

(d) As already mentioned, we consider a sequence of observation schemes which becomes more and more dense as n increases. In a good part of the paper, we consider equidistant sampling: at stage n we have a mesh Δ_n and sampling occurs at all times

$i\Delta_n$ for $i = 0, 1, \ldots$, within $[0, T]$. However, we also occasionally consider the case where the sampling times are possibly random, hence at stage n the number of observations within $[0, T]$ is random as well.

One reason for being interested in the volatility is that it constitutes one of the key building blocks for a model such as (1), and in the continuous case with $X_t^J \equiv 0$ this is indeed the only part of the model that can be estimated in a consistent way as $n \to \infty$. In finance, another reason is that spot volatility is a fundamental ingredient of hedging strategies, and is also heavily used for pricing. So the literature for this problem is really huge, mainly in econometrics, and it is out of the question in this short review to explain all the methods which have been proposed. We will focus on the main, or more popular, ones, and proofs are omitted.

Before starting, we give some notation and conventions. The sampling times are denoted, at stage n, by $T(n, i)$ for $i \geq 0$, with the convention $T(n, 0) = 0$. The $T(n, i)$'s are possibly random but finite-valued and strictly increasing in i, and for simplicity of notation we allow them to be bigger than T, but the statistics that are introduced should not use $X_{T(n,i)}$ when $T(n, i) > T$. We also denote as $N_t^n = \sum_{i\geq 1} 1_{\{T(n,i)\leq t\}}$ the number of observed returns, up to any time t. The ith return of X is denoted as $\Delta_i^n X = X_{T(n,i)} - X_{T(n,i-1)}$. The regular sampling case corresponds to having $T(n, i) = i\Delta_n$ for some sequence Δ_n of positive numbers, decreasing to 0, and in this case $N_t^n = [t/\Delta_n]$ (with $[x]$ the integer part of the real x).

2 The continuous case

In this section, we consider (1) in the continuous case $X^J \equiv 0$, that is

$$dX_t = b_t\,dt + \sigma_t\,dW_t. \tag{2}$$

This is defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and b_t, σ_t are two optional processes, respectively locally integrable and locally square-integrable.

Under (2), C_t is the quadratic variation of the process X at time t, and the approximate quadratic variation (also called *realized volatility*, or *realized variance*)

$$\widehat{C}_t^n = \sum_{i=1}^{N_t^n} (\Delta_i^n X)^2 \tag{3}$$

converges to C_t in probability as soon as the mesh of the observation grid goes to 0 and each $T(n, i)$ is a stopping time, without any additional assumption on the two processes b_t and c_t. We even have

$$\widehat{C}^n \overset{\text{u.c.p.}}{\Longrightarrow} C, \quad \text{where } \overset{\text{u.c.p.}}{\Longrightarrow} \text{ is convergence in probability, uniform on each interval } [0, t]. \tag{4}$$

Therefore the $\widehat{C}_t^n$'s form a (weakly) consistent sequence of estimators of C_t, and the question of its rate of convergence naturally arises. In other words, do the estimators $\widehat{C}_t^n$ enjoy a Central Limit Theorem (CLT), in the sense that $Z_t^n = w_n(\widehat{C}_t^n - C_t)$

converges (for a given t, or as processes indexed by t) to a non-trivial limit for some sequence w_n of real going to ∞?

The CLTs below always take the same form: after centering and normalization, a sequence Z^n of variables or processes defined on $(\Omega, \mathcal{F}, \mathbb{P})$ converges *stably in law* to a limit Z. This means two things:

1. The variable Z is defined on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ of the space $(\Omega, \mathcal{F}, \mathbb{P})$, meaning that $\widetilde{\Omega} = \Omega \times \Omega'$ for an auxiliary space Ω' (so any set or function or σ-field on Ω can be considered, with the same symbol, as a set or function or σ-field on $\widetilde{\Omega}$), and $\widetilde{\Omega}$ is equipped with a σ-field $\widetilde{\mathcal{F}}$ and a probability measure $\widetilde{\mathbb{P}}$ such that $\mathcal{F} \subset \widetilde{\mathcal{F}}$ and $\mathbb{P}(A) = \widetilde{\mathbb{P}}(A)$ when $A \in \mathcal{F}$.
2. Assuming that Z^n and Z take their values in some Polish space E, for any bounded continuous function f on E and any bounded $\mathcal{F}$-measurable variable V one has

$$\mathbb{E}\big(f(Y_n)\,V\big) \;\to\; \widetilde{\mathbb{E}}\big(f(Y)\,V\big).$$

We write $Z^n \stackrel{\mathcal{L}-s}{\longrightarrow} Z$ for this convergence, or $Z^n \stackrel{\mathcal{L}-s}{\Longrightarrow} Z$ when Z^n and Z are càdlàg (= right continuous with left hand limits) processes and E is the Skorokhod space equipped with the J_1 topology. That $Z^n \stackrel{\mathcal{L}-s}{\longrightarrow} Z$ implies $Z^n \stackrel{\mathcal{L}}{\longrightarrow} Z$ is obvious, and stable convergence in law has two additional properties not shared by mere convergence in law. First

$$\text{if } Z^n \stackrel{\mathcal{L}}{\longrightarrow} Z \text{ and } Y^n \stackrel{\mathbb{P}}{\longrightarrow} Y \text{ on } (\Omega, \mathcal{F}, \mathbb{P}), \text{ then } \quad (Y^n, Z^n) \stackrel{\mathcal{L}-s}{\longrightarrow} (Y, Z). \tag{5}$$

This property, used for studentizing estimators, is the reason for which Renyi (1963) introduced this type of convergence. The second property is that, unlike convergence in law, stable convergence in law makes sense also *restriction to a set* $A \in \mathcal{F}$:

$$Z^n \stackrel{\mathcal{L}}{\longrightarrow} Z \;\text{ in restriction to } A \text{ if (5) holds for all } f \text{ and all } V \text{ vanishing outside } A. \tag{6}$$

The limits encountered below are always as follows, depending on whether one considers variables or processes:

1. The variable Z is, conditionally on $\mathcal{F}$, a centered normal variable with some variance Σ (an $\mathcal{F}$-measurable variable); we write $Z^n \stackrel{\mathcal{L}-s}{\longrightarrow} MN(0, \Sigma)$, or $Z^n \stackrel{\mathcal{L}-s}{\longrightarrow} \mathcal{N}(0, \Sigma)$ when Σ is not random.
2. The process Z_t is, conditionally on $\mathcal{F}$, a centered Gaussian martingale, whose $\mathcal{F}$-conditional law is thus fully determined by the process $\Sigma_t = \widetilde{\mathbb{E}}((Z_t)^2 \mid \mathcal{F})$, and we write $Z^n \stackrel{\mathcal{L}-s}{\Longrightarrow} MGM(0, \Sigma)$.

2.1 Regular sampling

Here we suppose that $T(n,i) = i\Delta_n$. Let us begin with the simple case where $b_t \equiv b$ and $c_t \equiv c$ are two constants, so (2) is the Black–Scholes model for a log-price. The returns $\Delta_i^n X$ are i.i.d. as i varies, with the law $\mathcal{N}(b\Delta_n, c\Delta_n)$, so the

$\widehat{C}_t^n$'s are partial sums of a random walk, and with proper centering and normalization Donsker's theorem immediately yields

$$\frac{1}{\sqrt{\Delta_n}}\,(\widehat{C}_t^n - C_t) \stackrel{\mathcal{L}}{\Longrightarrow} \sqrt{2}\,cB_t$$

where $\stackrel{\mathcal{L}}{\Longrightarrow}$ stands for the functional convergence in law of process (for Skorokhod J_1 topology), and B is another Brownian motion. In particular, when $t = 1$ and $\Delta_n = 1/n$, we have $\sqrt{n}(\widehat{C}_1^n - c) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, 2c^2)$. Moreover the sequence of estimators $\widehat{C}_t^n$ is asymptotically efficient for estimating $C_t = ct$ in LeCam's sense (not surprising indeed: we really are in a classical parametric setting here, the LAN property holds, and $\widehat{C}_t^n$ is actually the MLE when $b = 0$ and a quasi-MLE otherwise).

Coming back to the general situation of (2), we do have the following theorem:

Theorem 1 *If the two processes b_t and c_t are locally square-integrable, we have the (functional) stable convergence in law* $\frac{1}{\sqrt{\Delta_n}}\,(\widehat{C}_t^n - C_t) \stackrel{\mathcal{L}-s}{\Longrightarrow} MGM(0, Q)$, *where Q_t is the process*

$$Q_t = 2\int_0^t (c_s)^2\,\mathrm{d}s. \tag{7}$$

Using the property (6), we deduce the following convergence for any $t > 0$ and any weakly consistent sequence $\widetilde{Q}_t^n$ of estimators for Q_t:

$$\frac{1}{\sqrt{\Delta_n\,\widetilde{Q}_t^n}}\,(\widehat{C}_t^n - C_t) \stackrel{\mathcal{L}-s}{\longrightarrow} \mathcal{N}(0,1) \quad \text{in restriction to the set } \{C_t > 0\} \tag{8}$$

(functional convergence in t is totally excluded here). In practice, one typically assumes that $C_t > 0$ a.s., so we have the usual convergence in law to a standard normal variable. (8) and standard techniques allow one to construct confidence regions for C_t, exactly as if it were an ordinary (non-random) statistical parameter. A sequence of estimators $\widetilde{Q}_t^n$ satisfying the previous conditions is

$$\widetilde{Q}_t^n = \frac{2}{3\Delta_n}\sum_{i=1}^{N_t^n}(\Delta_i^n X)^4. \tag{9}$$

How good are the estimators $\widehat{C}_t^n$? A general answer to this question is so far unknown, and indeed the various definitions of asymptotic efficiency, such as the LAN or LAMN property or the Hajek convolution theorem, do not apply to the case of arbitrary processes (subject to the condition of the theorem) $b_t,\ c_t$. However, Clément et al. (2013) have shown the following, as soon as the process σ_t is itself a continuous Itô semimartingale driven by a Brownian motion independent of W (no leverage), plus some mild technical conditions: a convolution theorem holds, asserting that for any sequence $\widehat{C}_t^{\prime n}$ of estimators such that $u_n(\widehat{C}_t^{\prime n} - C_t) \stackrel{\mathcal{L}-s}{\longrightarrow} Z_t'$ for some limit Z_t', we have $\limsup_n u_n\sqrt{\Delta_n} < \infty$ and when $u_n = 1/\sqrt{\Delta_n}$ then Z_t' can be written as

$Z'_t = Z_t + \overline{Z}_t$, where $\overline{Z}_t$ is $\mathcal{F}$-conditionally independent of Z_t. In order words, in those cases $\widehat{C}^n_t$ is asymptotically efficient. And, it is common practice to assert, in the general case, that a sequence of estimators satisfying the conclusions of the previous theorem is asymptotically efficient.

2.2 Irregular sampling

Now we consider (possibly) irregular sampling, and write $\Delta(n,i) = T(n,i) - T(n,i-1)$. Recall first that each $T(n,i)$ should be a stopping time, that the mesh $\rho_n = \max(\Delta(n,i) : i = 1, \ldots, N^n_T)$ should go to 0, at least in probability, and that $T(n,i)$ increases a.s. to ∞ as $i \to \infty$. The consistency (4) still holds, but for the CLT things are more complicated and necessitate quite strong assumptions.

The first assumption is about the structure of the $T(n,i)$'s, which should be more than mere stopping times. It basically says that $T(n,i)$ may depend on the past of X up to $T(n,i-1)$, plus possibly some exogenous random input, but is independent of the path of X on the interval $(T(n,i-1), T(n,i)]$; below, F^X_t denotes the σ-field generated by all variables X_t.

Assumption (A) Conditionally on $\mathcal{F}_{T(n,i-1)}$, the variable $T(n,i)$ is independent of $\mathcal{F}^X$.

This is of course satisfied by any irregular but deterministic sampling scheme, and also by what can be called a "modulated random walk" scheme where

$$\Delta(n,i) = \lambda_{T(n,i-1)}\, \Phi^n_i\, \Delta_n \tag{10}$$

where Δ_n is a sequence of positive numbers going to 0 and λ_t is a positive optional process and, for each n, the $(\Phi^n_i;\ i \geq 1)$ are i.i.d. positive variables with mean 1, independent of $\mathcal{F}^X$. This includes Poisson schemes where each process N^n_t is Poisson with intensity $1/\Delta_n$ and independent of $\mathcal{F}^X$.

This assumption makes sense in a financial context: the $T(n,i)$'s are typically transaction times, and an agent decides to buy or sell on the basis of what is known about X up to the previous transaction, plus perhaps some personal random preferences. However, this assumption excludes some interesting cases, such as when the $T(n,i)$ are successive hitting times of a spatial grid by X.

The second assumption is more technical. It depends on a real $q \geq 0$ and runs as follows:

Assumption (B-q) There are a sequence Δ_n of positive numbers going to 0 and an optional process $a(q)$ such that, for all t,

$$\Delta_n^{1-q} \sum_{i=1}^{N^n_t} \Delta(n,i)^q \ \xrightarrow{\mathbb{P}}\ \int_0^t a(q)_s \,\mathrm{d}s.$$

(B-1) holds always, with $a(1)_t = 1$. A regular sampling scheme with mesh size Δ_n satisfies (B-q) with $a(q)_t = 1$ and the same Δ_n. The modulated random walk scheme (10) satisfies (B-q) with the same Δ_n and $a(q)_t = (\lambda_t)^{q-1}\mu_q$ as soon as $\mathbb{E}((\Phi^n_i)^q) \to \mu_q$ and $1/\lambda_t$ is locally bounded.

Theorem 2 *If b_t is locally bounded and c_t is càdlàg and* (*A*) *and* (*B*-2) *hold, we have* $\frac{1}{\sqrt{\Delta_n}}\,(\widehat{C}^n_t - C_t) \stackrel{\mathcal{L}-s}{\Longrightarrow} MGM(0, Q'_t)$, *where*

$$Q'_t = 2\int_0^t (c_s)^2\, a(2)_s\, \mathrm{d}s.$$

In the genuine irregular case, Δ_n is a mathematical abstraction entering (B-2) but is not observable: if we replace Δ_n by $\alpha\Delta_n$ for some $\alpha > 0$, (B-2) still holds with $a(2)_t$ substituted with $a(2)_t/\alpha$. So, for a feasible CLT of the type (8) we need to consistently estimate Q'_t and also get rid of Δ_n. Under the additional assumption (B-4), the variables $\widetilde{Q}^n_t$ of (9) are consistent estimators for Q'_t, so (8) still holds. And, since $\Delta_n\,\widetilde{Q}^n_t = \sum_{i=1}^{N^n_t}(\Delta^n_i X)^4$ is observable, this constitutes a truly feasible result.

In the previous setting, one can refer to Ait-Sahalia and Mykland (2003) and Hayashi et al. (2011). We should also mention the following fact: whereas, as mentioned already, most results expounded in this paper extend to the multivariate case quite straightforwardly (except for notation…), irregular sampling schemes bring forth specific problems in the multivariate case because the sampling scheme may differ from one component of the process to the others (the so-called asynchronous observations case). A mere application of the previous method would lead to the so-called Epps effect, which results in a systematic downward bias for estimating the integrated absolute value of the quadratic covariation between any two components. To alleviate this effect, Hayashi and Yoshida (2008) and several subsequent papers have introduced a quite original method, which unfortunately cannot be explained here, due to the lack of space.

3 When there are jumps

In this section, we consider regular sampling schemes only.

When the process X of (1) has jumps, the situation is radically different. Indeed, as soon as it is a semimartingale (a property which is deemed as necessary for a price or log-price process because of the first fundamental asset pricing theorem), the estimators $\widehat{C}^n_t$ converge in probability to the quadratic variation $[X, X]_t$, which is

$$[X, X]_t = C_t + \sum_{s\le t}[\Delta X_s)^2, \tag{11}$$

where $\Delta X_s = X_s - X_{s-}$ is the size of the jump at time s, with the convention $\Delta X_0 = 0$. The sum above has at most countably many nonzero entries, and is a.s. finite for all t. So, as soon as there is at least one jump within $(0, t]$, the estimators $\widehat{C}^n_t$ are no longer consistent for C_t.

For constructing consistent estimators, we need some assumptions on the jump part X^J. Since the continuous part is assumed to be a continuous Itô semimartingale, it is only natural to assume that X^J is a purely discontinuous Itô semimartingale. By this, one means two properties:

1. We have $X_t^J = \sum_{s\leq t} \Delta X_s\, 1_{\{|\Delta X_s|>1\}} + M_t$, with M a local martingale without continuous part.
2. For any Borel set A of $\mathbb{R}$ at a positive distance of 0, the counting process $\mathcal{N}_t^A = \sum_{s\leq t} 1_A(\Delta X_s)$ has a predictable compensator which is absolutely continuous with respect to Lebesgue measure. Equivalently, we have a collection $F_{\omega,t}(dx)$ of measures on $\mathbb{R}$ (called "spot Lévy measures") such that, for any A as above, the process $F_t(A)$ is optional and

$$\mathcal{N}_t^A - \int_0^t F_s(A)\,\mathrm{d}s \quad \text{is a local martingale.}$$

   Necessarily, in this case the process $\int (x^2 \wedge 1)\, F_t(dx)$ is locally integrable.

Another notion is important for us. With the convention $0^0 = 0$, for any $r \geq 0$ we set

$$D(r)_t = \sum_{s\leq t} \big(|\Delta X_s|^r \wedge 1\big), \qquad \widetilde{D}(r)_t = \int_0^t \int ((|x|^r \wedge 1)\, F_s(\mathrm{d}x)\,\mathrm{d}s$$

(so $D(0)_t$ is the number of jumps on $[0,t]$). These processes are non-decreasing in t and non-increasing in r. They are a.s. finite-valued when $r \geq 2$, and when $r \leq 2$ then $D(r)$ is a.s. finite-valued if and only if $\widetilde{D}(r)$ is such. The number

$$\begin{aligned}\beta &= \inf\big(r > 0:\ D(r)_t < \infty \ \text{ for all } t \in \mathbb{R}_+\big)\\ &= \inf\big(r > 0:\ \widetilde{D}(r)_t < \infty \ \text{ for all } t \in \mathbb{R}_+\big)\end{aligned}$$

is called the *Blumenthal–Getoor index*, or BG index for short) of the process X, and necessarily $\beta \in [0,2]$. Notice that if $r < \beta$ we have $\mathbb{P}(D(r)_t = \infty) > 0$ for some t, whereas $D(\beta)_t$ can be finite-valued or not, depending on the case, and the same for $\widetilde{D}(r)$.

With all these notions at hand, we come back to the estimation of the integrated volatility C_t. Several methods have been proposed in the literature, and we emphasize four of them, the first two ones being the most popular ones.

3.1 Truncated variances

This method has been proposed by Mancini (2001, 2011) and consists in using (3) after deleting the "big" returns. More specifically, we choose a sequence v_n of positive numbers, and set

$$\widehat{C}(v_n)_t^n = \sum_{i=1}^{N_t^n} (\Delta_i^n X)^2\, 1_{\{|\Delta_i^n X|\leq v_n\}}. \tag{12}$$

When $v_n \equiv v > 0$ and if X has no jump of absolute size v, then $\sum_{i=1}^{N_t^n} (\Delta_i^n X)^2\, 1_{\{|\Delta X|>v\}}$ is easily seen to converge (pathwise, for all t) to $\sum_{s\leq t} (\Delta X_s)^2\, 1_{\{|\Delta X_s|>v\}}$. Compared with (11) and since $\widehat{C}_t^n \xrightarrow{\mathbb{P}} [X,X]_t$, we deduce that $\widehat{C}(v_n)_t^n \xrightarrow{\mathbb{P}} C_t +$

$\sum_{s\le t}(\Delta X_s)^2\, 1_{\{|\Delta X_s|\le v\}}$. If we let $v_n \to 0$ it is then natural to expect that $\widehat{C}(v_n)^n_t \xrightarrow{\mathbb{P}} C_t$, under appropriate conditions: if v_n goes too fast to 0 we could have $\widehat{C}(v_n)^n_t = 0$, and if it goes too slowly we would not eliminate the small jumps when the BG index is large.

We need some assumptions, with $r \in [0, 2]$ (recall that a *localizing sequence* is a sequence of stopping times increasing a.s. to infinity).

Assumption (H-r) X is an Itô semimartingale, σ is càdlàg, and there are a localizing sequence τ_n and constants Γ_n such that

$$t < \tau_n(\omega) \;\;\Rightarrow\;\; |b_t(\omega)[+|\sigma_t(\omega)| + \int (|x|^r \wedge 1)\, F_{\omega,t}(\mathrm{d}x) \le \Gamma_n.$$

(H-r) implies that $D(r)$ and $\widetilde{D}(r)$ are finite-valued, so for instance (H-0) implies the so-called finite activity for jumps, and (H-1) implies that the jumps of X are locally summable. Note also that (H-r) implies $\beta \le r$, and (H-r) implies (H-r') for all $r' \ge r$.

Below we take the truncation levels as follows:

$$v_n \asymp \Delta_n^{\varpi} \qquad \text{for some } \varpi \in (0, 1/2), \tag{13}$$

where for two sequences a_n, b_n of positive numbers we write $a_n \asymp b_n$ if the two sequences a_n/b_n and b_n/a_n are bounded.

Under (H-2) and (13), the consistency $\widehat{C}(v_n)^n_t \xrightarrow{\mathbb{P}} C_t$ holds for all t, and this convergence in probability is even uniform in $t \in [0, T]$ for any finite T. The associated CLT is:

Theorem 3 *Assume (H-r) for some $r \in [0, 1)$ and take $\varpi \in \left(\frac{1}{4-2r}, \frac{1}{2}\right)$ in (13). Then we have $\sqrt{\Delta_n}\left(\widehat{C}(v_n)^n_t - C_t\right) \stackrel{\mathcal{L}-s}{\Longrightarrow} MGM(0, Q_t)$, with Q_t as in (7).*

Then $\widehat{C}(v_n)^n_t$ satisfies (8), if we replace $\widetilde{Q}^n_t$ by the following estimators for Q_t (since, as for $\widehat{C}^n_t$, the estimators $\widetilde{Q}^n_t$ are no longer consistent):

$$\widetilde{Q}(v_n)^n_t = \frac{2}{3\Delta_n}\sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)^4\, 1_{\{|\Delta_i^n X| \le v_n\}}.$$

We thus see that the estimators $\widehat{C}(v_n)^n_t$ are asymptotically efficient, but only under (H-r) for some $r < 1$, and the tuning parameters v_n depend on this value r. In (12), we could be more specific and impose the form $v_n = \alpha\Delta_n^{\varpi}$ for some constant $\alpha > 0$. In practice, r is typically unknown (except if we use a model with finite activity, so (H-0) holds), but even if it were known we could take ϖ in the above range but we would still have to choose the constant α. However, in real life practitioners typically use a rule of thumb, according for instance to the discussion of Sect. 6.2.2 of Ait-Sahalia and Jacod (2014).

Now, one can wonder whether the restriction $r < 1$ in this theorem is a drawback of the particular truncated estimators $\widehat{C}(v_n)^n_t$. The answer is a no, and this restriction is in some sense an intrinsic limitation, due to the nonparametric nature of the problem.

To understand this, we can look at the class $\mathcal{S}^r$ of all semimartingales satisfying (H-r), each $X \in \mathcal{S}^r$ being defined on its own probability space. We consider any "universal" estimator (for each frequency Δ_n) which estimates the integrated volatility $C(X)_t$ for all $X \in \mathcal{S}^r$, so this estimator takes the form $f_t^n(X)$ for a suitable function f_t^n on $\mathcal{S}^r$ (this is obviously the case of the truncated estimators above).

We say that this sequence of estimators achieves the *uniform rate* w_n over some subclass $\mathcal{S}'^r \subset \mathcal{S}^r$ if the family of variables $w_n(f_t^n(X) - C(X)_t)$ is tight, uniformly in n and $X \in \mathcal{S}'^r$. Of course uniform rates cannot exist on $\mathcal{S}^r$ itself and we have to restrict our attention to subsets $\mathcal{S}'^r$ which are bounded in some sense. Toward this aim, it seems natural to consider the following set, for any $r \in [0, 2]$ and $\Gamma > 0$ and η-integrable function J on E:

$$\mathcal{S}_\Gamma^r \;=\; \text{the set of all semimartingales satisfying (H-}r\text{) with } \tau_1 \equiv \infty \text{ and } \Gamma_1 = \Gamma.$$

Then the following is proved in Jacod and Reiss (2013):

Theorem 4 *Let* $r \in [0, 2)$.

(a) *Any uniform rate* w_n *for estimating* $C(X)_t$ *within the class* $\mathcal{S}_\Gamma^r$ *satisfies*

$$w_n \le \begin{cases} 1/\sqrt{\Delta_n} & \text{if } r \le 1 \\ \left(\frac{\log(1/\Delta_n)}{\Delta_n}\right)^{(2-r)/2} & \text{if } r > 1. \end{cases} \tag{14}$$

(b) *Taking* $v_n = \Delta_n^{\varpi}$, *the truncated estimators* $\widehat{C}(v_n)_t^n$ *have the following uniform rate within the class* $\mathcal{S}_\Gamma^r$:

$$w_n = \begin{cases} 1/\sqrt{\Delta_n} & \text{if } r < 1 \text{ and } \frac{1}{4-2r} \le \varpi < \frac{1}{2} \\ 1/\Delta_n^{\varpi(2-r)} & \text{if } r \ge 1 \text{ and } 0 < \varpi < \frac{1}{2}. \end{cases}$$

Of course, this is a typical nonparametric result and it might happen that, for a particular $X \in \mathcal{S}^r$ with $r > 1$, some estimators exist which converge with the rate $1/\sqrt{\Delta_n}$ (we will actually encounter this in the third method below). However, it enlightens the limitation $r < 1$ in Theorem 3, and shows that it cannot be relaxed without additional structural hypotheses on the process X.

3.2 Multipower variations

The second method has been introduced by Barndorff-Nielsen and Shephard, plus Winkel for the CLT (Barndorff-Nielsen and Shephard 2004; Barndorff-Nielsen et al. 2006). We choose some integer $k \ge 2$ and set, with m_p being the pth absolute moment of an $\mathcal{N}(0, 1)$ variable for any $p \ge 0$, the estimators

$$\widehat{C}_t^{k,n} = \frac{1}{(m_{2/k})^k} \sum_{i=0}^{[t/\Delta_n]-k} \prod_{j=1}^{k} |\Delta_{i+j}^n X|^{2/k}.$$

One could take in the above product different positive powers, provided they add up to 2, but an equi-powers product seems natural. For the Black–Scholes model, the property $\widetilde{C}_t^{k,n} \xrightarrow{\mathbb{P}} C_t$ is essentially trivial. This property also holds when X is continuous, and when X has jumps as well because there cannot be two "big" jumps in the same interval $(i\Delta_n, (i+k)\Delta_n]$ when Δ_n is small, hence in the products the jumps are kind of wiped out.

The assumptions needed here are quite stronger than (H-r). First, the process σ_t itself should be an Itô semimartingale, which without restriction can be written as

$$d\sigma_t = b_t^\sigma\, dt + H_t^\sigma\, dW_t + H_t'^\sigma\, dW_t' + d\sigma_t^J, \tag{15}$$

with the same W as in (1) and another independent Brownian motion W', suitable coefficients b^σ, H^σ, H'^σ, and σ^J is the purely discontinuous part of σ (as in (1) for X).

Assumption (K-r) We have (H-r) for X and (H-2) for σ, and b_t in (1) and H_t^σ in (15) are càdlàg.

Here again (K-r) becomes stronger when r decreases, and the best available result is as follows (see Vetter (2010) for the fact that when $k=2$ the same result does not hold):

Theorem 5 *Assume (K-2/3) and also that σ_t and σ_{t-} never vanish. Then $\frac{1}{\sqrt{\Delta_n}}(\widehat{C}_t^{3,n} - C_t) \stackrel{\mathcal{L}-s}{\Longrightarrow} MGM(0, Q_t'')$, where*

$$Q_t'' = \alpha \int_0^t (c_s)^2\, ds, \quad \text{with} \ \ \alpha = \frac{(m_{4/3})^4 + (m_{4/3})^2 (m_{2/3})^2 - 7 m_{4/3} (m_{2/3})^6 + 5 (m_{2/3})^8}{(m_{4/3})^6 \big((m_{4/3} - (m_{2/3})^2\big)}.$$

Under the conditions of this theorem, the estimators $\widehat{C}_t^{3,n}$ satisfy (8), if we replace $\widetilde{Q}_t^n$ by

$$\widetilde{Q}_t''^n = \frac{\alpha}{(m_{4/3})^3\, \Delta_n} \sum_{i=0}^{[t/\Delta_n]-3} \prod_{j=1}^{3} |\Delta_{i+j}^n X|^{4/3}.$$

The advantage of this method is that it needs no tuning parameter. The drawback is that it requires stronger assumptions on X (in particular (H-r) for $r = 2/3$), and the estimators are rate efficient, but not variance efficient.

3.3 Local empirical characteristic functions

The third method we expound takes its origin in Jacod and Todorov (2014), and the proofs are still unpublished (they are in a working paper form, by the same authors). It is more complicated than the previous ones, and unfortunately requires two distinct tuning parameters. On the other hand, under appropriate assumptions it allows us to obtain asymptotically efficient estimators for C_t for an arbitrary value of the BG index in $[0, 2)$. It is based on the empirical characteristic functions of the returns, computed for an argument u_n going to ∞ within local windows of ν_n successive returns, and then aggregated over all windows after taking the logarithm.

We choose a sequence of reals $u_n \geq 1$ and a sequence of integers $\nu_n \geq 1$, satisfying

$$u_n \asymp \frac{1}{\ell_n \sqrt{\Delta_n}}, \qquad \nu_n \asymp \frac{1}{(\ell_n)^5 \sqrt{\Delta_n}}, \qquad \text{where } \ \ell_n = \log(1/\Delta_n). \tag{16}$$

For any $y > 0$, we consider the real and imaginary parts of the empirical characteristic function over a window starting at $i\Delta_n$:

$$L^{\mathfrak{R}}(y)_i^n = \frac{1}{\nu_n} \sum_{j=1}^{\nu_n} \cos(y u_n \Delta_{i+j}^n X), \qquad L^{\mathfrak{I}}(y)_i^n = \frac{1}{\nu_n} \sum_{j=1}^{\nu_n} \sin(y u_n \Delta_{i+j}^n X).$$

With the convention $\log 0 = 0$, we set

$$L(y)_i^n = -\log\left(|L^{\mathfrak{R}}(y)_i^n|^2 + |L^{\mathfrak{I}}(y)_i^n|^2\right),$$

and the aggregated logarithmic transform process is

$$\mathfrak{L}(y)_t^n = \frac{\nu_n}{u_n^2} \sum_{i=0}^{[t/\nu_n\Delta_n]-1} \left(L(y)_{i\nu_n}^n - \frac{1}{\nu_n}\left(e^{-L(y)_{i\nu_n}^n} - 1\right)\right).$$

Why do we use such a complicated statistics? To see that, consider the Black–Scholes case for which the $\Delta_i^n X$ are i.i.d. with law $\mathcal{N}(b\Delta_n, c\Delta_n)$. If ν_n is large, in a kind of heuristic way we have

$$L^{\mathfrak{R}}(y)_i^n \approx e^{-y^2 u_n^2 \Delta_n c/2} \cos(y u_n b \Delta_n), \qquad L^{\mathfrak{I}}(y)_i^n \approx e^{-y^2 u_n^2 \Delta_n c/2} \sin(y u_n b \Delta_n),$$

so $|L^{\mathfrak{R}}(y)_i^n|^2 + |L^{\mathfrak{I}}(y)_i^n|^2 \approx e^{-y^2 u_n^2 \Delta_n c}$ and $L(y)_i^n \approx y^2 u_n^2 \Delta_n c$, so finally $\mathfrak{L}(y)_t^n \approx y^2 ct$ and we may expect $\mathfrak{L}(1)_t^n$ to be consistent estimators of $ct = C_t$ (we do not need local windows in this case).

Analogously, in the case of (2), and provided b_t and σ_t are "smooth enough" in t, we have $L(y)_i^n \approx y^2 u_n^2 \Delta_n c_{i\Delta_n}$, implying $\mathfrak{L}(y)_t^n \approx y^2 C_t$ (we do need local windows here).

When there are jumps, the heuristic approach above still works, except that we now have an additional term coming from the jumps. This additional term turns out to be negligible under (K-r) with $r < 1$, plus the following mild assumption on the drift term, for some localizing sequence τ_m:

$$\mathbb{E}((b_{(t+s)\wedge\tau_m} - b_{t\wedge\tau_m})^2 \mid \mathcal{F}_t) \leq \Gamma_m\, s \qquad \text{for some constant } \Gamma_m, \tag{17}$$

and we have:

Theorem 6 *Under (K-r) for some $r \in [0, 1)$ and (17), we have $\frac{1}{\sqrt{\Delta_n}}\left(\mathfrak{L}(1)_t^n - C_t\right) \stackrel{\mathcal{L}-s}{\Longrightarrow}$ MGM$(0, Q_t)$.*

This result in itself has no real interest, since it requires slightly stronger assumptions than Theorem 3 and two tuning parameters. The interest of the method is that, up to a mild modification, it works also when (K-r) holds for some $r \geq 1$ under a special structure of the jumps, which can be formulated as follows, in a somewhat loose way: the jump part is the sum of a purely discontinuous Itô semimartingale satisfying (H-r) for some $r < 1$ plus a stochastic integral with respect to a purely discontinuous Lévy process which is "close enough" to a possibly asymmetric stable process with index $\beta \in (0, 2)$, so the BG index of X is β when $\beta > r$, and can thus be arbitrary in [0, 2).

The proper formulation is as follows. The process X satisfy

$$dX_t = b_t\,dt + \sigma_t\,dW_t + \sigma'_{t-}\,dY_t + dX'^J_t, \tag{18}$$

under the following assumption:

Assumption (S) (a) The processes σ and σ' are Itô semimartingales satisfying (H-2), and both b_t and H^σ_t (recall (15)) satisfy (17).

(b) The process X'^J is a purely discontinuous Itô semimartingales satisfying (H-r) for some $r < 1$.

(c) The process Y is a purely discontinuous Lévy process (no drift and no Gaussian component) whose Lévy measure F satisfies for some nonnegative constants a^+, a^-, K and some $r' \in [0, 1 \wedge \beta)$ and all $x > 0$:

$$\left|F((x,\infty)) - \frac{a^+}{x^\beta}\right| + \left|F((-\infty,-x)) - \frac{a^-}{x^\beta}\right| \leq \frac{K}{x^{r'}}. \tag{19}$$

For example, a tempered stable process with index β satisfies (19).

Coming back to our heuristic approach, it can be shown that under Assumption (S) we have

$$\mathfrak{L}(y)^n_t \approx y^2 C_t + u_n^{\beta-2}\chi_\beta y^\beta A_t, \quad \text{where } A_t = \int_0^t (a^+_s + a^-_s)\,\mathrm{d}s \tag{20}$$

with χ_β an (explicit) number depending only on β, so we have a bias term of order $u_n^{\beta-2}$ which, under (16), is not negligible in front of the rate $1/\sqrt{\Delta_n}$ when $\beta \geq 1$. However, a simple de-biasing procedure consists in taking, for some $y > 1$ and with the convention $0/0 = 0$, the estimators

$$\widehat{C}'(y)^n_t = \mathfrak{L}(1)^n_t - \frac{\left(\mathfrak{L}(y)^n_t - y^2\mathfrak{L}(1)^n_t\right)^2}{\mathfrak{L}(y^2)^n_t - 2y^2\mathfrak{L}(y)^n_t + y^4\mathfrak{L}(1)^n_t}. \tag{21}$$

Theorem 7 *Under (S) and for any $t > 0$, we have $\frac{1}{\sqrt{\Delta_n}}\,(\widehat{C}'(y)^n_t - C_t) \xrightarrow{\mathcal{L}-s} MN(0, Q_t)$.*

This means that the sequence $\widehat{C}'(y)^n_t$ is again asymptotically efficient for estimating C_t. When $\beta \geq 1$, this seems in contradiction with Theorem 4, which tells us that the best possible rate should satisfy the second part of (14) for any $r > \beta$, instead of being the much faster $1/\sqrt{\Delta_n}$. However, that theorem is about the uniform rate over the whole class $\mathcal{S}^r$ of Itô semimartingales, whereas here we consider a much smaller class, with a rather specific structure for the jumps: so here we really are in a semi-parametric setting instead of the fully nonparametric situation of Theorem 4, and there is no mathematical contradiction.

One should add a few comments about Theorem 7:

(1) The result holds for any given t, and could be extended as to have a joint convergence for a finite family of times $0 < t_1 < \cdots < t_k$, but there cannot be a functional convergence in t.

(2) When $\beta \geq 1$ the result holds only on the set $\{A_t > 0, C_t > 0\}$, on which both the Brownian part and the stable-like jump part have been active on a subset of $[0, t]$ with positive Lebesgue measure. The behavior of the estimators is unknown outside this set.

(3) (S) could be significantly weakened, with X an Itô semimartingale of the form (1) and its spot Lévy measures F_t satisfying (after localization) the property (18), with a_t^+ and a_t^- two processes satisfying (17), plus some (mild) extra conditions. However, in practice models with high activity jumps are mostly of the form (18), or as described in the next comment, so for simplicity we stick to this case.

(4) One can consider a model analogous to (18), with $\sigma'_{t-}dY_t$ replaced by a sum $\sum_{m=1}^{M} \sigma'^m_{t-} dY^m_t$, where the Y^m's are independent processes of the same type as Y above, with distinct indices β_m. In the approximate expression (20), we then have M bias terms. One can iterate the de-biasing procedure (21) to get again asymptotically efficient estimators.

3.4 Nearest neighbor truncation

This method has been introduced by Andersen et al. (2012) and consists in taking one of the following two estimators, where $\mathrm{med}(x, y, z)$ stands for the number among x, y, z which is between the other two values;

$$\widehat{C}_t^{\min,n} = \frac{\pi}{\pi - 2} \sum_{i=1}^{N_t^n - 1} \min((\Delta_i^n X)^2, (\Delta_{i+1}^n X)^2)$$

$$\widehat{C}_t^{\mathrm{med},n} = \frac{\pi}{6 - 4\sqrt{3} + \pi} \sum_{i=1}^{N_t^n - 2} \mathrm{med}((\Delta_i^n X)^2, (\Delta_{i+1}^n X)^2, (\Delta_{i+2}^n X)^2)$$

(the factors in front of the sums are the inverses of $\mathbb{E}(\min(U_1, U_2))$ and $\mathbb{E}(\mathrm{med}(U_1, U_2, U_3))$, for U_1, U_2, U_3 independent standard normal variables). By the same kind of argument as for multipowers, this methods eliminates the jumps.

Theorem 8 *Assume* (K-0) *and also that σ_t and σ_{t-} never vanish. Then* $\frac{1}{\sqrt{\Delta_n}}(\widehat{C}_t^{min,n} - C_t) \stackrel{\mathcal{L}-s}{\Longrightarrow} MGM(\alpha_{min}\, Q_t)$ *and* $\frac{1}{\sqrt{\Delta_n}}(\widehat{C}_t^{med,n} - C_t) \stackrel{\mathcal{L}-s}{\Longrightarrow} MGM(\alpha_{med}\, Q_t)$*, where* $\alpha_{\min} \approx 1.9$ *and* $\alpha_{\mathrm{med}} \approx 1.48$.

We get rate efficiency, but not quite variance efficiency. The asymptotic variance for $\widehat{C}_t^{\min,n}$ and a fortiori the one for $\widehat{C}_t^{\min,n}$ are in fact smaller than the asymptotic variance in Theorem 5. So this method is preferable to the multipower variation method, but *only under* (K-0) (implying finite activity of jumps). It is likely that it works under (K-r) for $r < 1$ as well, but this has not been formally proven so far.

4 Estimating the spot volatility

Here again we suppose that the sampling scheme is regular. The aim is to estimate the spot volatility c_t. One can pose this problem in two different ways:
(1) One may want to estimate c_t at a given time t, for hedging purposes for example. In this case, we use observations around t, and typically those in a relatively small time window right before t (for hedging we need to devise a strategy at time t, using data prior to t).
(2) One may want to "reconstruct" the whole path of the process $t \mapsto c_t$ for $t \in [0, T]$, for the particular sample point ω which has been (partially) observed: so we need all observed data between 0 and T.

One might (correctly) argue that solving (1) for all $t \in [0, T]$ automatically yields a solution of (2). However, (1) is a "local" problem solved by using a local method, and the final criterion is the estimation error at time t. On the other hand, (2) is a "global" problem for which the final criterion is the distance (in $L^p([0, T])$ for some $p = 1$ or $p = 2$ or $p = \infty$, for example) between the estimator and its target, and the estimation error at a particular time t does not matter much. This difference is the same as in between pointwise and functional estimations of a density for an i.i.d. sequence.

4.1 Local estimation

The idea is indeed very simple. For each t, take any sequence $\overline{C}_t^n$ of consistent estimators for C_t, so $\overline{C}_{t+s}^n - \overline{C}_t^n \xrightarrow{\mathbb{P}} C_{t+s} - C_t$ if $s, t > 0$ and use the fact that $\frac{1}{s}\int_t^{t+s} c_r \,\mathrm{d}r$ converges to c_t as $s \to 0$, as soon as c is right continuous at t, so we may hope that

$$\frac{1}{s_n}\left(\overline{C}_{t+s_n}^n - \overline{C}_t^n\right) \xrightarrow{\mathbb{P}} c_t$$

when $s_n \to 0$, and in practice, s_n should be an integer multiple of Δ_n, say $s_n = w_n \Delta_n$ for $w_n \geq 1$.

For $\overline{C}_t^n$, we will either use $\widehat{C}_t^n$ of (3) when X is continuous, or $\widehat{C}(v_n)_t^n$ of (12) when X has jumps. This leads us to set, with any sequence v_n satisfying $v_n \asymp \Delta_n^{\varpi}$ for some $\varpi \in (0, 1/2)$:

$$\begin{array}{ll} \text{Non-truncated estimator:} & \widehat{c}(w_n)_i^n = \frac{1}{w_n\Delta_n}\sum\limits_{j=1}^{w_n}(\Delta_{i+j}^n X)^2 \\ \text{Truncated estimator:} & \widehat{c}(w_n, v_n)_i^n = \frac{1}{w_n\Delta_n}\sum\limits_{j=1}^{w_n}(\Delta_{i+j}^n X)^2\, 1_{\{|\Delta_{i+j}^n X| \leq v_n\}}. \end{array} \tag{22}$$

We always assume (at least) that c_t is càdlàg, so for any given t the estimators for c_{t-} and c_t will be, respectively, and according the case where X is continuous or not,

$$\left.\begin{array}{ll} \widehat{c}_{t-}(w_n)^n = \widehat{c}(w_n)^n_{i-w_n-1}, & \widehat{c}_{t-}(w_n, v_n)^n = \widehat{c}(w_n, v_n)^n_{i-w_n-1} \\ \widehat{c}_t(w_n)^n = \widehat{c}(w_n)^n_i, & \widehat{c}_t(w_n, v_n)^n = \widehat{c}(w_n, v_n)^n_i \end{array}\right\} \text{ if } (i-1)\Delta_n < t \le i\Delta_n,$$
$$\widehat{c}_0(w_n)^n = \widehat{c}(w_n)^n_0.$$

The first estimators do not anticipate on what happens after time t, whereas the second ones do, and actually if c jumps exactly at time t we need to anticipate if we want to estimate the value c_t after the jump. However, typically the process c has no fixed times of discontinuity (when it is an Itô semimartingale, for example), so for any fixed t we have $c_t = c_{t-}$ and can use either one of the two estimators, or a mixture of them. On the other hand, we can in principle replace t above by a stopping time, which may be a jump time of c, so then one should be careful about whether one uses the "left" or "right" estimators.

The consistency of these estimators is ensured under quite minimal assumptions

Theorem 9 *Assume* (*H*-2) *and take* w_n *such that* $w_n \to \infty$ *and* $w_n\Delta_n \to 0$. *Then* $\widehat{c}_t(w_n)^n \xrightarrow{\mathbb{P}} c_t$ *for all* $t \ge 0$ *and* $\widehat{c}_{t-}(w_n)^n \xrightarrow{\mathbb{P}} c_{t-}$ *for all* $t > 0$, *and the same for the truncated estimators under* (13).

This result is remarkable in the sense that even when X has jumps we do not need to use the truncated versions. This is due to the fact, when t is fixed, there is (with a probability going to 1) no big jump in the interval $[t - w_n\Delta_n, t + w_n\Delta_n]$.

When it comes to the associated CLT, though, we do need the truncated versions when X jumps. We also need a more specific behavior for the sequence w_n, namely

$$w_n\sqrt{\Delta_n} \to \alpha \quad \text{for some } \alpha \in [0,\infty], \qquad w_n \to \infty, \qquad w_n\Delta_n \to 0. \tag{23}$$

We also need (15), which implies that c has a form similar to (15) and thus a volatility c^c_t given by

$$c^c_t = c_t((H^\sigma_t)^2 + (H'^\sigma_t)^2).$$

Assuming (23), the CLT is as follows:

Theorem 10 (a) *If X satisfies* (*K*-*r*) *for some* $r \in [0, 4/3)$ (*hence in particular when it satisfies* (*K*-2) *and is continuous*) *and under* (23), *for any* $t \ge 0$ *we have*

$$\sqrt{w_n}\left(\widehat{c}_t(w_n)^n - c_t\right) \xrightarrow{\mathcal{L}-s} MN(0, 2(c_t)^2) \qquad \text{if } \alpha = 0 \tag{24}$$

$$\sqrt{w_n}\left(\widehat{c}_t(w_n)^n - c_t\right) \xrightarrow{\mathcal{L}-s} MN(0, 2(c_t)^2 + \alpha^2 c^c_t/3) \qquad \text{if } 0 < \alpha < \infty \tag{25}$$

$$\frac{1}{\sqrt{w_n\Delta_n}}\left(\widehat{c}_t(w_n)^n - c_t\right) \xrightarrow{\mathcal{L}-s} MN(0, \alpha^2 c^c_t/3) \qquad \text{if } \alpha = \infty, \tag{26}$$

and the same for $\widehat{c}_{t-}(w_n)^n$ *when* $t > 0$. *The same holds for the truncated estimators under* (13).[1]

(b) *When* X *satisfies* (*K-r*) *for some* $r \in [4/3, 2)$, *and if* $w_n \asymp \Delta_n^{-\tau}$ *and* $v_n \asymp \Delta_n^{\varpi}$, *the truncated estimators* $\widehat{c}_t(w_n, v_n)^n$ *and* $\widehat{c}_{t-}(w_n, v_n)^n$ *enjoy the convergence* (24) *when* $\tau \in \left(0, \frac{2-r}{2}\right)$ *and* $\varpi \in \left(\frac{\tau}{2(2-r)}, \frac{1}{2}\right)$, *and the convergence* (26) *when* $\tau \in \left(2-\frac{2}{r}, 1\right)$ *and* $\varpi \in \left(\frac{1-\tau}{2(2-r)}, \frac{1}{2}\right)$.

Some comments are in order here.

(1) The estimation error in the case of $\widehat{c}(w_n)_i^n$ when X is continuous, say, admits the decomposition

$$\begin{aligned}\widehat{c}(w_n)_i^n - c_{i\Delta_n} = & \frac{1}{w_n\Delta_n}\left(\widehat{C}^n_{(i+w_n)\Delta_n} - \widehat{C}^n_{i\Delta_n}) - (C_{(i+w_n)\Delta_n} - C_{i\Delta_n}\right) \\ & + \frac{1}{w_n\Delta_n}\int_{i\Delta_n}^{(i+w_n)}(c_s - c_{i\Delta_n})\,\mathrm{d}s,\end{aligned}$$

we have two competing terms, as usual in nonparametric estimation; the first error term above is (statistically) bigger than the second one when $\alpha = 0$ in (23), smaller when $\alpha = \infty$, and both have the same order of magnitude when $\alpha \in (0, \infty)$; this explains the three different regimes found in (a) of the theorem, and of course the best rate $1/\Delta_n^{1/4}$ is obtained when $\alpha \in (0, \infty)$.

(2) When X has jumps the rate-optimal convergence (25) holds under (K-r), when $r < 1$ as in Theorem 4, but also when $1 \le r < 4/3$, which is again quite remarkable. When $r \ge 4/3$, though, we only have sub-optimal rates.

(3) Considering again (24) for simplicity (but the following holds for the other two convergences as well), we do have a multidimensional version

$$\left(\sqrt{w_n}\left(\widehat{c}_t(w_n)^n - c_t\right), \sqrt{w_n}\left(\widehat{c}_{t-}(w_n)^n - c_t\right)\right)_{t\in\mathcal{T}} \xrightarrow{\mathcal{L}-s} \left((V_t, V'_t)\right)_{t\in\mathcal{T}}$$

for any *finite* set $\mathcal{T}$ of times and where, conditionally on $\mathcal{F}$, all variables V_t and V'_t are independent centered Gaussian, with the same variance $2(c_t)^2$ for V_t and V'_t. However, a functional (it t) is excluded, because the limit would be $\mathcal{F}$-conditionally a (non-homogeneous) white noise.

(4) About the feasibility of the estimators, meaning the possibility of finding consistent estimators for the asymptotic variances, allowing one to derive confidence bounds: in the case of (24) there is no problem, since $2(\widehat{c}_t(w_n)^n)^2$ are consistent estimators, and we readily deduce

$$\frac{\sqrt{w_n}}{\sqrt{2}\,\widehat{c}_t(w_n)^n}\left(\widehat{c}_t(w_n)^n - c_t\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{in restriction to the set } \{c_t > 0\}.$$

In the other cases, we need also consistent estimators for c_t^c. Those are in principle available, for example in Sect. 8.3 in Ait-Sahalia and Jacod (2014) estimators for

[1] This is an opportunity to correct a mistake in Jacod and Protter (2012): the factor $1/3$, which appears in (13.3.9) on page 391, should also appear in the last two formulas of the display (13.3.11). The same factor $1/3$ should also appear in the second part of (8.14), p. 269 of Ait-Sahalia and Jacod (2014).

the integrated version $\int_0^t c_s^c\,ds$ are constructed, with the rate $1/\Delta_n^{1/4}$, allowing (in principle) to get spot estimators for c_t^c converging with the rate $1/\Delta_n^{1/8}$ (under appropriate, rather strong, assumptions, and upon taking still another sequence of tuning parameters). The conclusion is that reliable estimation of c_t^c is not really available. Therefore, when estimating c_t we would rather advise to use a sequence w_n such that $w_n\sqrt{\Delta_n} \asymp 1/\log(1/\Delta_n)$ (so $\alpha = 0$), giving rise to estimators with sub-optimal rate $1/\Delta_n^{1/4}\sqrt{\log(1/\Delta_n}$ when (K-r) for some $r < 4/3$, but which are feasible.

5) Our last comment is about the choice of α in (23), say when X is continuous. The best rate is obtained when $\alpha \in (0,\infty)$, and actually $\Delta_n^{-1/4}(\widehat{c}_t(w_n)^n - c_t) \stackrel{\mathcal{L}-s}{\longrightarrow} MN(0, 2(c_t)^2/\alpha^2, c_t^c/3)$ in this case. So in principle one should choose α large to minimize the estimation error. In the same vein, instead of the choice (22) for $\widehat{c}(w_n)_i^n$ we could take a genuine kernel estimator of the form

$$\widehat{c}'(w_n)_i^n = \frac{1}{w_n\Delta_n}\sum_{j=1}^{w_n} f(j/w_n)\,(\Delta_{i+j}^n X)^2$$

for some smooth function f on $[0,1]$ with $\int_0^1 f(x)\,\mathrm{d}x = 1$, as usually done for pointwise nonparametric estimation. The optimal choice of f would be the same as if c_t had $1/2$–Hölder continuous paths (although this is not true, for example this process might jump), and is of course not the uniform kernel $f \equiv 1$ used in (22).

However, since we already advised the reader to rather choose a slightly sub-optimal procedure with $\alpha = 0$, we also advise to take the version (22) with the uniform kernel, since any other choice would simply increase the estimation variance which would be proportional to $\int_0^1 f(x)^2\,\mathrm{d}x$, which is bigger than 1 under the constraint $\int_0^1 f(x)\,\mathrm{d}x = 1$ unless $f \equiv 1$.

4.2 Global estimation and the Fourier method

Now we turn to the global estimation of spot volatility. As already mentioned, the previous spot estimators can also be considered as global estimator: either $\widehat{c}_t(w_n)^n$ or $\widehat{c}_{t-}(w_n)^n$ or $\frac{1}{2}(\widehat{c}_t(w_n)^n+\widehat{c}_{t-}(w_n)^n)$ do the job. Whereas there cannot exist a functional CLT, we have a rather good estimate of the supremum of the estimation error. Namely, if $w_n \asymp 1/\sqrt{\Delta_n}$, it has been proved by Fan and Wang (2008) that under (K-2) and when X is continuous and $H'^{\sigma} \equiv 0$ in (15) (no leverage) we have

$$\text{the sequence } \frac{1}{\Delta_n^{1/4}\sqrt{\log(1/\Delta_n)}}\sup_{t\in[0,T-w_n\Delta_n}|\widehat{c}_t(w_n^n - c_t| \text{ is tight}$$

and the same with $\widehat{c}_{t-}(w_n)^n$ or $\frac{1}{2}(\widehat{c}_t(w_n)^n + \widehat{c}_{t-}(w_n)^n)$. This result almost certainly holds also in the presence of leverage, and probably also when X jumps and (K-r) holds for some $r < 4/3$.

Another method, based on Fourier transforms, is also very powerful (it has been introduced by Malliavin and Mancino (2002) and subsequently developed under various conditions in Mancino and Sanfelici (2012), for example). As in all global methods, the horizon T plays a fundamental role. Below, since i is usually an index in this paper, we use the notation $\iota = \sqrt{-1}$. Unfortunately, so far it has really developed for the continuous case only, so below we assume that both X and c are continuous.

There is some arbitrariness in the definition of the Fourier coefficients and here, for any relative integer $k \in \mathbb{Z}$, the kth Fourier coefficient of a function g on $[0, T]$ is

$$\mathcal{F}_k(g) = \frac{1}{2\pi} \int_0^T e^{-2\iota\pi kt/T}\, g(t)\, \mathrm{d}t.$$

At stage n and for the function $g(t) = c_t$ on $[0, T]$, and with a suitable sequence of integers $w_n \to \infty$, it will be estimated by

$$\widehat{\mathcal{F}_k}(w_n)^n = \frac{1}{2w_n + 1} \sum_{r=-w_n}^{w_n} a_{-r}^n\, a_{r+k}^n, \qquad \text{with } a_r^n = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{N_T^n} e^{-2\iota\pi\, i\, r\Delta_n/T}\, \Delta_i^n X.$$

The rationale behind this is as follows: suppose that c_t is non-random and continuous and that X has no drift and is continuous. Since $\mathbb{E}(\Delta_i^n X\, \Delta_{i'}^n X)$ is approximately $c_{i\Delta}\Delta_n$ when $i' = i$ and vanishes otherwise, $\mathbb{E}(a_{-r}^n\, a_{r+k}^n)$ converges to $\mathcal{F}_k(c)$ by Riemann approximation. Although the variables $a_{-r}^n\, a_{r+k}^n$ are not independent as r varies, they enjoy a law of large numbers. Since $w_n \to \infty$, it is thus not surprising that

$$\widehat{\mathcal{F}_k}^n \xrightarrow{\mathbb{P}} \mathcal{F}_k(c).$$

When c_t is random, and in presence of a drift term, the argument is more complicated, but in any case the above consistency still holds.

We also have a CLT, under the additional assumption that the process σ_t (in the initial papers it was also assumed that c_t admits a Malliavin derivative with suitable moment bounds for all t, but this assumption was relaxed later on, see, e.g., Clément and Gloter 2013 or Cuchiero and Teichmann 2015). Then, if w_n satisfies $T/w_n\Delta_n \to \alpha \in (0, \infty)$, the sequence $\frac{1}{\sqrt{\Delta_n}}\,(\widehat{\mathcal{F}_k}(w_n)^n - \mathcal{F}_k(c))$ converges stably in law to a limit which is $\mathcal{F}$–conditionally centered Gaussian variable with conditional variance $u(\alpha, k) \int_0^T (c_s)^2\, \mathrm{d}s$, with $u(\alpha, k)$ suitable constants depending on k and α only, and we do have indeed a joint convergence when k varies.

In particular, since C_T is equal to $2\pi\mathcal{F}_0(c)$, this gives us an alternative method for estimating C_T, and indeed when $w_n = [T/2\Delta_n]$ (the so-called Nyquist frequency) the sequence of estimators $\widehat{\mathcal{F}_0}(w_n)/(2\pi)$ are asymptotically efficient for that purpose (but this is much more complicated than $\widehat{C}_T^n$ and needs stronger assumptions).

Coming back to spot estimation, the Fourier–Fejer formula allows us to reconstruct the path of c_t from its Fourier coefficients, leading to the estimators

$$\widehat{c}(T, w_n)_t^{\text{Fourier},n} = \frac{1}{2w_n + 1} \sum_{k=-w_n}^{w_n} \left(1 - \frac{|k|}{w_n}\right) e^{2\iota\pi kt/T}\, \widehat{\mathcal{F}_k}(w_n)^n$$

which is necessarily nonnegative. It enjoys the following uniform consistency property, as soon as $w_n\Delta_n \to 0$:

$$\sup_{t\in[0,T]} |\widehat{c}(T,w_n)_t^{\mathrm{Fourier},n} - c_t| \xrightarrow{\mathbb{P}} 0.$$

As far as a CLT is concerned, in the literature it is restricted to targets of the form $\int_0^T g(t)c_t\,\mathrm{d}t$ for any reasonable non-random function g (in Malliavin and Mancino (2011) or Clément and Gloter (2013) for example), or $\int_0^T g(c_t)\,\mathrm{d}t$ for smooth enough functions g (in Cuchiero and Teichmann (2015) for example): since this is somehow outside our main topic we will not state the results. An interesting but still open question (as far as I know) would be to find a good rate v_n, depending on w_n and Δ_n, such that

$$v_n^p \int_0^T |\widehat{c}(T,w_n)_t^{\mathrm{Fourier},n} - c_t|^p\,\mathrm{d}t \qquad \text{is bounded in probability.}$$

Concerning robustness of the estimators against jumps of X, virtually nothing is known (except for the estimation of $\int_0^T g(c_t)\,\mathrm{d}t$ under strong growth conditions on g, excluding the case $g(x)=x$ for example): the estimators have to be modified, of course, using perhaps a truncation method, but how exactly is not clear.

On the other hand, the Fourier method works when the sampling is irregular (under simple and natural conditions, when the $T(n,i)$'s are non-random, at least), and in the multivariate case for asynchronous sampling. This can be viewed as an alternative method to the Hayashi–Yoshida method. The Fourier method is also robust against noise, at least in the white noise case.

5 When there is microstructure noise in the continuous case

This section is devoted to the case where the observations are contaminated with a microstructure noise. That is, at each observation time $T(n,i)$ we do not exactly observe $X_{T(n,i)}$ but a noisy version

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

The observed returns are denoted as $\Delta_i^n Y = Y_i^n - Y_{i-1}^n$, and we assume that X is continuous.

5.1 Structure of the noise

The noisy situation has been the object of quite many papers, mostly when the noise is a *white noise*, meaning that for each n the variables ε_i^n are centered i.i.d. (Gaussian or not, but typically with finite moments of all orders) and are furthermore independent of the process X. Two different situations have indeed been considered in the white noise setting:

1. *Constant noise case* by this, one means that the laws of the variables ε_i^n do not depend on n, or at least their variances do not depend on n.
2. *Shrinking noise case* the variances $\gamma_n = \mathbb{E}((\varepsilon_i^n)^2)$ go to 0 as $n \to \infty$ (the case $\gamma_n \to \infty$ of an exploding noise clearly has no practical relevance). If the convergence $\gamma_n \to 0$ is fast enough, namely $\gamma_n/\Delta_n^{3/2} \to 0$, the noise is asymptotically negligible and all results in the previous section hold if we replace everywhere $X_{T(n,i)}$ by Y_i^n (if we only have $\gamma_n/\Delta_n \to 0$ the consistency results still hold, but not the CLTs). In contrast, when the convergence is slower, the previous estimators are subject to an exploding bias.

In finance, the price may be real-valued (hopefully positive), but the observations are based on transaction prices, which are multiples of a minimal value, typically 1 cent. As a consequence, the order of magnitude of the observation error is also 1 and the observed returns at high frequency (say with a 1 second time lag) are quite often 0 or 1; that means that noise is bigger than average returns. In other words, a model with a shrinking noise is not adapted to this situation, and below we restrict our attention to the case of a "constant" noise.

Another feature, which again follows from the previous comment, is that the observations Y_i^n take their values on a grid with mesh 1 if they are prices, or on the corresponding "log grid" if they are log-prices. Therefore, the noise ε_i^n cannot be independent of X since it is constrained by the fact that $X_{T(n,i)} + \varepsilon_i^n$ should be an integer, or a log-integer. Thus a white noise model, although it gives us a useful intuition for the noisy case, is highly not adapted to an asset price process.

Therefore, we devote some time below to develop a noise model which covers situations where the observations Y_i^n are necessarily integers or log-integers, and at the same time allows us to conduct reasonable estimation procedures for the volatility.

The first idea which comes to one's mind is to take for Y_i^n the rounded value of $X_{T(n,i)}$, that is $Y_i^n = [X_{T(n,i)}]$ (this could also be, probably more accurately, $[X_{T(n,i)} + 1/2]$, which is the integer closest to $X_{T(n,i)}$; we argue here for prices, but the same would hold for log-prices, with more cumbersome notation). It might very well be that this is the genuine structure of the noise, but in this case we are in a bad shape: if the full path of X where observed after rounding, we would know the passage times (or, the local times) at each integer level, but nothing else, so C_t would *not* be known, and in the discrete observation case there cannot exist consistent estimators for C_t.

This leads us to introduce an assumption which allows for consistent estimation of C_t and which accommodates some form of rounding. Recall that the sampling times are $(\mathcal{F}_t)$-stopping times, and to accommodate for extra random input in the noise we suppose, without restriction, that the σ–field $\mathcal{F}_\infty$ is strictly included into $\mathcal{F}$.

Assumption (N) There is a càdlàg adapted process γ_t, a localizing sequence τ_m of stopping times such that the stopped processes $\gamma_{t\wedge\tau_n}$ satisfy (17) and that, for each n, conditionally on $\mathcal{F}_\infty$, the variables $(\varepsilon_i^n)_{i\geq 0}$ are independent and satisfy for all Borel subsets B of $\mathbb{R}$ and all $p > 0$:

$$\mathbb{E}(\varepsilon_i^n \mid \mathcal{F}_\infty) = 0, \qquad \mathbb{E}((\varepsilon_i^n)^2 \mid \mathcal{F}_\infty) = \gamma_{T(n,i)}$$
$$\mathbb{P}(\varepsilon_i^n \in B \mid \mathcal{F}_\infty) = \mathbb{P}(\varepsilon_i^n \in B \mid \mathcal{F}_{T(n,i)})$$

$$T_i^n < \tau_m \quad \Rightarrow \quad \mathbb{E}(|\varepsilon_i^n|^p \mid \mathcal{F}_\infty) \leq \Gamma(p)_m \quad \text{for some constants } Ga(p)_m. \quad (27)$$

The second line of (27) means that the noise at time $T(n,i)$ is not anticipating on the future. Of course, a white noise with finite moments satisfies (N), with $\gamma_t(\omega) = \gamma$ a constant.

Example 11 We give an example for which the observation is the rounded value of the process X plus a noise, such that (N) holds. Many versions are possible, and we describe the simplest one. Take a sequence α_i of i.i.d. variables, uniform on [0, 1] and independent of $\mathcal{F}_\infty$. The observations are

$$Y_i^n = \left[X_{T(n,i)} + \alpha_i\right].$$

Then one can show that (N) is satisfied. This example can be interpreted as follows: Y_i^n is a "randomized" version of $X_{T_i^n}$ which, on the set $\{m \leq X_{T_i^n} < m+1\}$, takes the value m with the probability $m+1-X_{T_i^n}$ and the values $m+1$ with the probability $X_{T_i^n} - m$, so Y_i^n is $[X_{T_i^n}]$ with a large probability when $X_{T_i^n}$ is "close" to its rounded value, and a small probability when $X_{T_i^n}$ is "far" from it, and $Y_i^n = [X_{T_i^n}]+1$ otherwise: this seems quite a reasonable (although somewhat simplistic) observation model.

Similar but more sophisticated models also satisfy (N): for example, we could replace the uniform law of α_i by a law which has the density $f(x) = a_m$ for $x \in (-m, -m+1] \cup (m, m+1]$ for any integer $m \geq 0$ (so $a_m \geq 0$ and $a_0 + 2\sum_{m\geq 1} a_m = 1$, plus $\sum_{m\geq 1} a_m m^p < \infty$ for all $p > 0$ to ensure the existence of moments).

It is also relatively simple to modify this example in order to fit a log-price instead of a price, with Y_i^n taking its values in the set $\{\log m : m = 1, 2, \ldots\}$.

A last remark, before proceeding: Assumption (N) is indeed a sophisticated version of a white noise, but in practice one might also have a "colored" noise, for which the ε_i^n are no longer independent or conditionally independent as i varies, hence the noise autocovariance is not vanishing. There is empirical evidence that this could really be the case for financial time series. However, with the exception of the pre-averaging method below, all de-noising methods so far are designed for noise with vanishing autocovariance, and we will not consider below the colored noise case.

5.2 Some preliminary remarks

The simplest—and least efficient—way to get rid of the noise is *subsampling*. Indeed, as seen before, the noise is troublesome only when its variance cannot be neglected in front of $\Delta_n^{3/2}$ (or rather, of $\overline{c}\Delta_n^{3/2}$, where $\overline{c}$ is the average volatility over the interval of interest).

For regular sampling, this basically means that, when Δ_n is small enough, $\widehat{C}_t^n - C_t$ is of order of magnitude $\overline{c}\sqrt{\Delta_n}$ if there is no noise, and in the presence of noise there is an additional error of order of magnitude γ/Δ_n, with γ the noise variance. Hence if $\gamma \ll \overline{c}\Delta_n^{3/2}$ the noise can really be neglected. For typical asset prices, this seems to be the case when Δ_n is at least a couple of minutes. When Δ_n is smaller than that, one can subsample the data, using only the observations at the times $im\Delta_n$ for some m

with $m\Delta_n$ bigger than, say, 5mn, and we can ignore the noise. However, this method uses only a small proportion of all available data, resulting in a huge loss of efficiency.

To circumvent this drawback, when X is continuous a large variety of methods have been proposed, mostly for white noise, and we briefly review some of them. However, it is enlightening to first recall the efficient rate at which C_t can hopefully be estimated, in the presence of a (constant) noise.

To this effect, consider the Black–Scholes model with a Gaussian white noise with variance γ, in the regular sampling case. The LAN property holds with the rate $1/\Delta_n^{1/4}$ for the parameter c and $1/\sqrt{\Delta_n}$ for the parameter γ, and the efficient asymptotic variance for estimating $C_t = ct$ is $8\gamma^{1/2}c^{3/2}t$. So, exactly as when there is no noise, in the case c_t is time-varying and (N) holds we call a sequence $\overline{C}_t^n$ of estimators of C_t *asymptotically efficient* (in the regular sampling case) if

$$\Delta_n^{-1/4}(\overline{C}_t^n - C_t) \xrightarrow{\mathcal{L}-s} MN\left(0, 8\int_0^t (\gamma_s)^{1/2}\,(c_s)^{3/2}\,\mathrm{d}s\right). \tag{28}$$

Another comment is in order here. With the exception of the quasi-MLE, all estimators $\overline{C}_t^n$ exhibited below can produce negative values, and actually do so with a positive probability in most cases. This is in deep contrast with the estimators in the previous sections when there is no noise. This can be considered as a serious drawback, but there really is nothing one can do about it, except for replacing $\overline{C}_t^n$ by 0 when it happens to be negative. We can also add the following remarks:

1. It might happen that $C_t = 0$, or that C_t is positive but very small; in these cases having a negative value of $\overline{C}_t^n$ is not overwhelmingly strange. When $C_t = 0$ for example, the probability that $\overline{C}_t^n > 0$ is typically close to 1/2.
2. Usually the estimator is used to derive a confidence interval; one really has to worry only when this interval is totally (or perhaps mostly) included into $(-\infty, 0)$.
3. Finding $\overline{C}_t^n < 0$ may also be an indication that either the model is incorrect and/or the assumptions violated, or that Δ_n is not small enough for the estimator to be close to the asymptotic regime.

We now proceed to a brief description of some of the methods introduced in the literature. Unless stated otherwise, the observation scheme is regular. A particular estimator $\overline{C}_t^n$ is called "local" if $\overline{C}_{t+s}^n - \overline{C}_t^n$ is equal (up to a negligible term in front of $\Delta_n^{1/4}$) to the same estimator applied to the data within the time interval $[t, t+s]$, and otherwise it is called "global." Local estimators can be used to estimate the spot volatility in the same way as in Sect. 4.1, whereas global ones typically cannot.

5.3 Quasi-maximum likelihood estimators (QMLE)

This method has been introduced (Ait-Sahalia et al. 2003) and the CLT proved by Xiu (2010), and it is about the continuous model (3) with a Gaussian white noise with variance γ. Below, we fix t.

In the Black–Scholes case, the MLE is of course asymptotically efficient, and the log-likelihood has an explicit form, which is as follows: for each pair $x, y > 0$ consider the $N_t^n \times N_t^n$ matrix with entries

$$\Sigma_n(x,y)^{i,j} = \begin{cases} x\Delta_n + 2y & \text{if } i = j \\ -y & \text{if } i = j \pm 1 \\ 0 & \text{otherwise,} \end{cases}$$

This matrix has an explicit inverse, and the log-likelihood in the Black–Scholes case with Gaussian with noise is the following, taken for $x = c$ and $y = \gamma$:

$$L(x,y)_n = -\frac{1}{2}\left(\log\det(\Sigma_n(x,y) + N_t^n \log(2\pi) + \sum_{i,j=1}^{N_t^n} (\Sigma_n(x,y)^{-1})^{i,j}\, \Delta_i^n Y\, \Delta_j^n Y\right). \tag{29}$$

The idea is then to maximize this log-likelihood, and the estimator for C_t is defined as follows:

$$\widehat{C}_t^{\mathrm{QMLE},n} = x_{min} t, \qquad \text{where} \quad (x_{\min}, y_{\min}) = \mathrm{argmin}_{\mathbb{R}_+\times\mathbb{R}_+} L(.,.)_n$$

(note that y_{min} would be an estimator of γ).

Although this procedure uses the wrong likelihood, we still have:

Theorem 12 *Assume that X is continuous and satisfies (K-2) and that the process $1/\sigma_t$ is locally bounded and the noise is a white noise. Then we have $\frac{1}{\Delta_n^{1/4}}(\widehat{C}_t^{\mathrm{QMLE},n} - C_t) \xrightarrow{\mathcal{L}-s} MN(0, Q_t^{\mathrm{QMLE}})$, where*

$$Q_t^{\mathrm{QMLE}} = \frac{\sqrt{\gamma}}{\sqrt{t\,C_t}}\left(5t\int_0^t (c_s)^2\,\mathrm{d}s + 3(C_t)^2\right).$$

The variance above is necessarily at least as big as the one in (28) (with $\gamma_t = \gamma$) but the two are difficult to compare, except when $c_t \equiv c$ is a constant, in which case they agree (since the estimator is then the true MLE). Note that $\widehat{C}_t^{\mathrm{QMLE},n}$ is by construction a global estimator.

With a suitable modification of the definition (29), this theorem holds also in the case of irregular sampling, when it is a modulated random walk, as defined in (10). On the other hand, whether one can relax the white noise assumption and only assume (N) is unknown.

5.4 Multi-scales estimators (MSRV)

Those estimators have been introduced by Zhang (2006) and are an extension of the previously introduced two-scales estimators (Zhang et al. 2005) which are simpler but unfortunately have the rate of convergence $1/\Delta_n^{1/6}$, instead of the efficient rate $1/\Delta_n^{1/4}$. They are based on a C^3 function f on [0, 1] and a tuning sequence k_n satisfying

$$\int_0^1 f(s)\,\mathrm{d}s = 0, \qquad \int_0^1 s\,f(s)\,\mathrm{d}s = 1, \qquad k_n\sqrt{\Delta_n} \to \theta \in (0,\infty). \tag{30}$$

The estimator at stage n is

$$\widehat{C}_t^{\mathrm{MSRV},n} = \sum_{j=1}^{k_n} \alpha_j^n \sum_{i=j}^{N_t^n} \left(Y_i^n - Y_{i-j}^n\right)^2, \tag{31}$$

where

$$\alpha_j^n = \frac{1}{k_n^2} f\left(\frac{j}{k_n}\right) - \frac{1}{2k_n^3} f'\left(\frac{j}{k_n}\right) - \frac{1}{6k_n^3} (f(1) - f(0)) + \frac{1}{24k_n^4} (f''(1) - f''(0)).$$

We associate with f the following quantities:

$$\begin{aligned}
A(f) &= \frac{1}{3}\int_0^1 f(x)\,\mathrm{d}x \int_0^x y^2(3x-y)\,f(y)\,\mathrm{d}y \\
A'(f) &= \int_0^1 f(x)\,\mathrm{d}x \int_0^1 (x \wedge y)\,f(y)\,\mathrm{d}y \\
A''(f) &= \int_0^1 f(x)^2\,\mathrm{d}x.
\end{aligned}$$

Then, under the white noise assumption and with the notation $\Gamma = \mathbb{E}((\eta_i^n)^4) - \gamma^2$ (which does not depend on n, i, we have:

Theorem 13 *Assume that X is continuous and satisfies* (K-2) *and that the noise is a white noise. Under* (30)*, we have* $\frac{1}{\Delta_n^{1/4}}\,(\widehat{C}_t^{\mathrm{MSRV},n} - C_t) \xrightarrow{\mathcal{L}-s} MN(0, Q_t^{\mathrm{MSRV}})$*, where*

$$Q_t^{\mathrm{MSRV}} = \frac{4}{\theta}\int_0^t \left(A(f)c_s^2 + 2A'(f)c_s\gamma\theta^2 + A''(f)\gamma^2\theta^4\right) ds + 4\Gamma\theta^3 A''(f). \tag{32}$$

The last term in (32) is due to the border summands (when i in the second sum in (31) is close to 0 or to N_t^n). This term does not have an obvious extension when we replace the white noise assumption by (N), so it is unlikely that the theorem extends to that situation. However, this estimator can be extended to irregular sampling schemes under appropriate assumptions.

At first glance, these estimators might look as being local ones. But this is not true, because again of the last term in (32).

5.5 Flat-top realized Kernels

Barndorff-Nielsen, Hansen, Lunde and Shephard (in a series of papers, see, e.g., Barndorff-Nielsen et al. 2008) introduced another type of kernel estimators, with a C^3 kernel f and a tuning sequence k_n satisfying

$$f(0) = 1, \quad f(1) = f'(0) = f'(1) = 0, \qquad k_n\sqrt{\Delta_n} \to \theta \in (0,\infty). \tag{33}$$

The estimators constructed below are really estimators for $C_{t-k_n\Delta_n} - C_{k_n\Delta_n}$, although they use all data between 0 and t, but since the rate is $1/\Delta_n^{1/4}$ they will also estimate C_t with the same asymptotic properties. We set

$$K_t^n = \sum_{i=k_n-1}^{N_t^n-k_n+2} \left((\Delta_i^n Y)^2 + \sum_{j=1}^{k_n-2} f\left(\frac{j-1}{k_n}\right) (\Delta_i^n Y\, \Delta_{i+j}^n Y + \Delta_i^n Y\, \Delta_{i-j}^n Y) \right).$$

It turns out that border effects seriously affect those estimators, which are typically not even consistent, because of the special role played by the first and last ith summand above. To overcome this difficulty, one uses the so-called jittering procedure, which basically amounts to replace the (k_n-2)th and $N_t^n - k_n + 2$th observation by the following averages, for some integer $m_n \le k_n - 1$:

$$\overline{Y}_{start}^n = \frac{1}{m_n} \sum_{i=0}^{m_n-1} Y_{k_n-2-i}^n, \qquad \overline{Y}_{end}^n = \frac{1}{m_n} \sum_{i=0}^{m_n-1} Y_{N_t^n+i-k_n+2}^n,$$

and the estimators finally are

$\widehat{C}_t^{\mathrm{FT},n}$ is defined as K_t^n above, after replacing $Y_{k_n-1}^n$ and $Y_{N_t^n-k_n+2}^n$ by $\overline{Y}_{start}^n$ and $\overline{Y}_{end}^n$.

The "flat-top" qualifier refers to the fact that in each ith summand the weights of $\Delta_i^n Y\, \Delta_{i+m}^n Y$ are equal (to 1) when $m = -1, 0, 1$. We associate with f the numbers

$$\Phi(f) = \int_0^1 f(s)^2\, \mathrm{d}s, \qquad \Phi'(f) = \int_0^1 f'(s)^2\, \mathrm{d}s, \qquad \Phi''(f) = \int_0^1 f''(s)^2\, \mathrm{d}s.$$

Theorem 14 *Assume that X is continuous and satisfies (K-2) and that the noise is a white noise. Under (33) plus $m_n \to \infty$ with $m_n\sqrt{\Delta_n} \to 0$, we have $\frac{1}{\Delta_n^{1/4}} (\widehat{C}_t^{\mathrm{FT},n} - C_t) \stackrel{\mathcal{L}-s}{\longrightarrow} MN(0, Q_t^{\mathrm{FT}})$, where*

$$Q_t^{\mathrm{FT}} = 4 \int_0^t \left(\Phi(f) c_s^2 \theta^{-1} + 2\Phi'(f) c_s \gamma \theta + \Phi''(f) \gamma^2 \theta^3 \right) \mathrm{d}s. \tag{34}$$

Comparing (32) with (34), we see that the latter does not exhibit the additional term which shows up in the former, explaining why the estimators $\widehat{C}_t^{\mathrm{FT},n}$ are indeed local estimators. But of course $\widehat{C}_t^{\mathrm{MSRV},n}$ and $\widehat{C}_t^{\mathrm{FT},n}$ are not really comparable, since the kernel f does not play quite the same role in the two estimators.

The previous theorem is for a white noise. When the noise is still independent of X but presents some serial dependence, a proper CLT usually fails; however $\widehat{C}_t^{\mathrm{FT},n}$ still converges to C_t and the convergence rate can be $1/\Delta_n^{1/4}$ in some cases and is in general controlled by the mixing rate of the noise, and there is a CLT (with rate $1/\Delta_n^{1/4}$ but a variance bigger than above) when the noise is AR(1) for example. There are also some results when the noise depends on X, but is a very specific way which excludes the case of Assumption (N) in general.

Now, how to choose the kernel f? For this, we have to compare the asymptotic variance (34) with the efficient one given in (28), and below we restrict our attention to the base case of the Black–Scholes model. We need to minimize the ratio

$$\alpha(\theta, f, c, \gamma) = \frac{1}{2\theta}\left(\Phi(f)c^{1/2}\gamma^{-1/2} + 2\Phi'(f)c^{-1/2}\gamma^{1/2}\theta^2 + \Phi''(f)c^{-3/2}\gamma^{3/2}\theta^4\right)$$

(always bigger than 1) by appropriately choosing θ and f. For a given f, the minimum is reached for

$$\theta = \sqrt{\frac{c}{\gamma}}\,\overline{\Phi}(f) \qquad \text{where} \quad \overline{\Phi}(f) = \left(\frac{\sqrt{\Phi'(f)^2 + 3\Phi(f)\Phi''(f)} - \Phi'(f)}{3\Phi''(f)}\right)^{1/2},$$

and with this choice $\alpha(\theta, f, c, \gamma)$ becomes independent of c, γ, with the value

$$\alpha_0(f) = \frac{1}{2\overline{\Phi}(f)}\left(\Phi(f) + 2\Phi'(f)\overline{\Phi}(f)^2 + \Phi''(f)\overline{\Phi}(f)^4\right).$$

So one needs to minimize $\overline{\Phi}(f)$, subject to the constraints (33). One does not know if there is a kernel such that $\alpha_0(f) = 1$, but some explicit kernels give a value very close to 1, such as for example:

$$\alpha_0(f) \approx \begin{cases} 1.13 & \text{cubic kernel: } f(x) = 1 - 3x^2 + 2x^3 \\ 1.07 & \text{Parzen kernel: } f(x) = \begin{cases} 1 - 6x^2 + 6x^3 & \text{if } x \le 1/2 \\ 2(1-x)^3 & \text{if } x > 1/2 \end{cases} \\ 1.0025 & \text{Tukey–Hanning kernel of order 16: } f(x) = \left(\sin(\pi/2(1-x)^{16})\right)^2. \end{cases}$$

Therefore, the Tukey–Hanning kernel is very close to optimality. However, in practice, and when c_t and γ_t are varying with time, the main problem is the choice of the number θ; therefore all choices among the f's above are essentially equivalent from a practical viewpoint. Note that some adaptive procedure, along the same lines as what we describe in the next section, might perhaps be available for flat-top estimators as well, but this as not been done so far.

5.6 The pre-averaging method

This method has been introduced by Podolskij and Vetter (2009) and also in Jacod et al. (2009). In contrast with the previous sections, we do not assume a white noise, but (N) should be satisfied. We still assume regular sampling. We need a kernel f on $[0, 1]$ and a sequence k_n of integers satisfying, for some $\theta \in (0, \infty)$,

$$f \text{ is continuous, piecewise } C^2, \quad f(0) = f(1) = 0, \qquad k_n\sqrt{\Delta_n} = \theta + \mathrm{o}(\Delta_n^{1/4}). \tag{35}$$

With f and k_n, we associate the quantities

$$\overline{f}_n = \sum_{j=1}^{k_n} f\left(\frac{i}{k_n}\right)^2, \qquad \overline{f}'_n = \sum_{j=1}^{k_n}\left(f\left(\frac{j}{k_n}\right) - f\left(\frac{j-1}{k_n}\right)\right)^2, \tag{36}$$

and also (with f' the derivative of f, well defined outside a finite set):

$$\begin{aligned}
\psi(f) &= \frac{1}{\psi(f)^2}\int_0^1 f(x)^2\,\mathrm{d}x, \qquad \psi'(f) = \int_0^1 f'(x)^2\,\mathrm{d}x\\
\Psi(f) &= \frac{1}{\psi(f)^2}\int_0^1 \Big(\int_x^1 f(y-x)f(y)\,\mathrm{d}y\Big)^2\,\mathrm{d}x,\\
\Psi''(f) &= \int_0^1 \Big(\int_x^1 f'(y-x)f'(y)\,\mathrm{d}y\Big)^2\,\mathrm{d}x\\
\Psi'(f) &= \frac{1}{\psi(f)^2}\int_0^1 \Big(\int_x^1 f(y-x)f(y)\,\mathrm{d}y\Big)\Big(\int_x^1 f'(y-x)f'(y)\,\mathrm{d}y\Big)\,\mathrm{d}x,
\end{aligned}$$

so in particular $\overline{f}_n = k_n\psi(f) + \mathrm{O}(1)$ and $\overline{f}'_n = \psi'(f)/k_n + \mathrm{O}(1/k_n^2)$.

We compute from the observation Y_i^n the following variables, implicitly depending on f and k_n:

$$\begin{aligned}
\overline{Y}_i^n &= \sum_{j=1}^{k_n-1} f\left(\frac{j}{k_n}\right)\Delta_{i+j-1}^n Y = -\sum_{j=1}^{k_n}\left(f\left(\frac{j}{k_n}\right) - f\left(\frac{j-1}{k_n}\right)\right) Y_{i+j-2}^n\\
\widehat{Y}_i^n &= \sum_{j=1}^{k_n}\left(f\left(\frac{j}{k_n}\right) - f\left(\frac{j-1}{k_n}\right)\right)^2 (\Delta_{i+j-1}^n Y)^2.
\end{aligned} \tag{37}$$

The first $\overline{Y}_i^n$ is the pre-averaged ith observed return, the second $\widehat{Y}_i^n$ is approximately equal to $2\psi'(f)\gamma_{i\Delta_n}/k_n$ and will be used as de-biasing terms below.

The estimators for C_t are given by

$$\widehat{C}_t^{\mathrm{Preav},n} = \frac{1}{\overline{f}_n}\sum_{i=1}^{N_t^n - k_n + 1}\Big((\overline{Y}_i^n)^2 - \frac{1}{2}\widehat{Y}_i^n\Big).$$

Theorem 15 *Assume that X is continuous with b_t locally bounded and σ_t càdlàg, and that the noise satisfies* (N)*. Under* (35)*, we have* $\frac{1}{\Delta_n^{1/4}}(\widehat{C}_t^{\mathrm{Preav},n} - C_t) \xrightarrow{\mathcal{L}-s} MN(0, Q_t^{\mathrm{Preav}})$*, where*

$$Q_t^{\mathrm{Preav}} = 4\int_0^t \big(\Psi(f)c_s^2\theta^{-1} + 2\Psi'(f)c_s\gamma_s\theta + \Psi''(f)\gamma_s^2\theta^3\big)\,\mathrm{d}s. \tag{38}$$

The variance (38) is the same as in (34), upon substituting $\Phi(f)$, $\Phi'(f)$, $\Phi''(f)$ by $\Psi(f)$, $\Psi'(f)$, $\Psi''(f)$. Therefore, in the Black–Scholes case plus a constant noise variance $\gamma_t = \gamma$, optimizing θ and f can be done as follows: first, take

$$\theta = \sqrt{\frac{c}{\gamma}}\,\overline{\Psi}(f) \qquad \text{where} \quad \overline{\Psi}(f) = \left(\frac{\sqrt{\Psi'(f)^2 + 3\Psi(f)\Phi''(f)} - \Psi'(f)}{3\Psi''(f)}\right)^{1/2}, \tag{39}$$

then try to minimize

$$\alpha_0'(f) = \frac{1}{2\overline{\Psi}(f)}\big(\Psi(f) + 2\Psi'(f)\overline{\Psi}(f)^2 + \Phi''(f)\overline{\Psi}(f)^4\big).$$

So, we have the same problem as with the flat-top kernel estimators. The analogy is not fortuitous; indeed, if we consider the function $g(x) = \frac{1}{\psi(f)}\int_x^1 f(y-x)f(y)\,dy$, the function g satisfies (33) as soon as f satisfies (35), and we then have $\Phi(g)\Psi(f)$ and $\Phi'(g) = \Psi'(f)$ and $\Phi''(g) = \Psi''(f)$, so $\alpha_0'(f) = \alpha_0(g)$. However, it is not clear how to deduce f from g, when the latter is the cubic, or Parzen, or Tukey–Hanning kernel. In practice, one typically use the following:

$$\text{triangular kernel:} \quad f(x) = 2\big(x \wedge (1-x)\big), \quad \text{for which} \;\; \alpha_0'(f) \approx 1.07$$

which is also close enough to optimality, and for which the averaged values $\overline{Y}_i^n$ have a nice interpretation, at least when $k_n = 2k_n'$ is even:

$$\overline{Y}_i^n = \frac{1}{k_n'}\big(Y_{i+k_n'-1}^n + \cdots + Y_{i+2k_n'-2}^n\big) - \frac{1}{k_n'}\big(Y_{i-1}^n + \cdots + Y_{i+k_n'-2}^n\big).$$

In order to make this CLT feasible, we need consistent estimators for Q_t^{Preav}, which is the case, for example, of

$$\begin{aligned}\widetilde{Q}_t^{\mathrm{Preav},n} &= \frac{1}{k_n\sqrt{\Delta_n}\,\psi(f)^2}\sum_{i=1}^{N_t^n-k_n+1}\Big(\frac{4\Psi(f)}{3\psi(f)^2}\,(\overline{Y}_i^n)^4 + 4\Big(\frac{\Psi'(f)}{\psi(f)\psi'(f)} - \frac{\Psi(f)}{\psi(f)^2}\Big)(\overline{Y}_i^n)^2\,\widehat{Y}_i^n \\ &\quad + \Big(\frac{\Psi(f)}{\psi(f)^2} - \frac{2\Psi'(f)}{\psi(f)\psi'(-f)} + \frac{\Psi''(f)}{\psi'(f)^2}\Big)(\widehat{Y}_i^n)^2\Big),\end{aligned}$$

Then, similar with (8), one gets for each $t > 0$:

$$\frac{1}{\Delta_n^{1/4}\sqrt{\widetilde{Q}_t^{\mathrm{Preav},n}}}\,(\widehat{C}_t^{\mathrm{Preav},n} - C_t) \overset{\mathcal{L}-\mathrm{s}}{\longrightarrow} MN(0,1) \quad \text{in restriction to the set } \{C_t > 0\}. \tag{40}$$

We finally mention that the pre-averaging method has been extended in Jacod et al. (2018) to the case of a colored noise, however independent of X.

5.7 Adaptive pre-averaging

Let us come back to the (sub-) optimal choice of θ and f. In a sense, the discussion of the previous section settles the problem for f: we do not have an "explicit" optimal kernel f (if it exists at all), but the triangular kernel is nearly optimal.

For θ, or equivalently k_n as a function of Δ_n through (35), this is another matter. The choice (39) is optimal when c_t and γ_t are constant in time. Otherwise, one can choose k_n "locally in time," according to the adaptive procedure introduced in Jacod and Mykland (2015). It seems important for this method that the sampling be regular.

The aim is still to estimate C_t for a given $t > 0$, and the kernel f is fixed throughout. We need two sequences m_n, l_n of integers, subject to the conditions

$$l_n \asymp \frac{1}{\Delta_n^w} \text{ with } w \in \left(\frac{5}{6}, 1\right), \quad m_n \asymp \frac{1}{\sqrt{\Delta_n}}, \quad 4 \leq 2m_n < l_n < t/\Delta_n. \tag{41}$$

We split the data into consecutive blocks of data with size l_n each (except for the last one, with size between l_n and $2l_n - 1$). The number of blocks is $L_n = [t/l_n\Delta_n]$ and the jth block starts at time $S(n, j-1) = T(n, (j-1)l_n)$ and ends at time $T(n, jl_n)$ if $j < L_n$ and t if $j = L_n$. We will actually use the pre-averaged quantities $\overline{Y}_i^n$ or $\widehat{Y}_i^n$ of (37) and the numbers $\overline{f}_n$ or $\overline{f}'_n$ of (36) for different sequences k_n, according to the block in which they are, and to keep track of this we rather write them as $\overline{Y}(k_n)_i$ or $\widehat{Y}(k_n)_i$, and $\overline{f}(k_n)$ or $\overline{f}'(k_n)$.

The procedure is divided into two steps.

Step 1 Local estimation of c_t and γ_t. We do the adaptive procedure for the blocks $j = 2, \ldots, L_n$, and for this we use estimators for c_t and γ_t based on the second half of the previous block. More specifically, we set

$$\widehat{c}^n(j) = \frac{2}{l_n\Delta_n\,\overline{f}(m_n)} \sum_{i=(j-1)l_n-[l_n/2]-m_n}^{(j-1)l_n-m_n} \left((\overline{Y}(m_n)_i)^2 - \frac{1}{2}\widehat{Y}(m_n)_i\right)$$

$$\widehat{\gamma}^n(j) = \frac{1}{l_n} \sum_{i=(j-1)l_n-[l_n/2]}^{(j-1)l_n} (\Delta_i^n Y)^2,$$

and, in view of (39) and with the convention $0/0 = 0$,

$$\widehat{\theta}^n(j) = \overline{\Psi}(f)\sqrt{\frac{\widehat{c}^n(j)}{\widehat{\gamma}^n(j)}}, \qquad k_{n,j} = \begin{cases} m_n & \text{if } j = 1 \\ [2\bigvee(\frac{l_n}{2}\bigwedge\frac{\widehat{\theta}^n(j)}{\sqrt{\Delta_n}})] & \text{if } j \geq 2. \end{cases}$$

Step 2 Global estimation For each block j, we estimate the integrated volatility with the pre-averaging estimator of the previous section and the tuning parameter $k_{n,j}$. The last block has a length usually smaller than the others, so we single out the two cases by setting

$$\begin{array}{llll} 1 \leq j < L_n \Rightarrow & J(n,j) = jl_n, & a(n,j) = 1 \\ j = L_n \Rightarrow & J(n,j) = [t/\Delta_n] - k_{n,L_n} + 1, & a(n,j) = \frac{J(n,L_n) - J(n,L_n-1) + k_{n,L_n}}{J(n,L_n) - J(n,L_n-1)}. \end{array}$$

and the estimator is

$$\widehat{C}_t^{\mathrm{Preav-Ad},n} = \sum_{j=1}^{L_n} \frac{a(n,j)}{\overline{f}(k_{n,j})} \sum_{i=J(n,j-1)+1}^{J(n,j)} \left((\overline{Y}(k_{n,j})_i)^2 - \frac{1}{2}\,\widehat{Y}(k_{n,j})_i \right).$$

Although the numbers $k_{n,j}$ are random when $j \geq 2$, this is well defined, and we have:

Theorem 16 *Assume that X is continuous and satisfies (K-2) with further σ_t locally bounded away from 0; assume also that the noise satisfies (N) with a process γ_t being an Itô semimartingale satisfying (H-2) and locally bounded away from 0. Under (41), we have $\frac{1}{\Delta_n^{1/4}}\,(\widehat{C}_t^{\mathrm{Preav-Ad},n} - C_t) \xrightarrow{\mathcal{L}-s} MN(0, Q_t^{\mathrm{Preav-Ad}})$, where*

$$Q_t^{\mathrm{Preav-Ad}} = 8\alpha_0'(f) \int_0^t (\gamma_s)^{1/2}\,(c_s)^{3/2}\,\mathrm{d}s.$$

Hence, if we compare with (28), we see that these estimators are asymptotically rate efficient, and the loss of variance efficiency is measured by the constant $\alpha_0'(f)$, very close to 1 for a triangular kernel.

Finally, these estimators are feasible in the sense of (40) if we replace $\widetilde{Q}_t^{\mathrm{Preav},n}$ by the following consistent estimators for $Q_t^{\mathrm{Preav-Ad}}$:

$$\begin{aligned} \widetilde{Q}_t^{\mathrm{Preav-Ad},n} = \frac{1}{\sqrt{\Delta_n}\,\psi(f)^2} \sum_{j=1}^{L_n} \frac{a(n,j)}{k_{n,j}} \sum_{i=J(n,j-1)+1}^{J(n,j)} \Big(& \frac{4\Psi(f)}{3\psi(f)^2}\,(\overline{Y}(k_{n,i})_i)^4 \\ & + 4\Big(\frac{\Psi'(f)}{\psi(f)\psi'(f)} - \frac{\Psi(f)}{\psi(f)^2}\Big)(\overline{Y}(k_{n,i})_i)^2\,\widehat{Y}(k_{n,i})_i \\ & + \Big(\frac{\Psi(f)}{\psi(f)^2} - \frac{2\Psi'(f)}{\psi(f)\psi'(-f)} + \frac{\Psi''(f)}{\psi'(f)^2}\Big)(\widehat{Y}(k_{n,i})_i)^2\Big), \end{aligned}$$

5.8 Final comments

Quite many other methods have been proposed in the literature. First, as already mentioned, the Fourier method works when there is a white noise, although a precise CLT in that case has never been stated as such.

Second, there are other kernel estimators, actually mostly developed in the multivariate case with asynchronous observations: this means that typically the $T(n,i)$'s (which are different for different components) are random, for instance modulated random walks as in (10), or satisfying an assumption like (B-q)) for $q = 2$ at least (interestingly enough, Assumption (A) is usually not necessary because the kernel kind of wipes out the possible dependence upon X, so some kinds of hitting times may be eligible).

Let us mention for example Bibinger et al. (2014) for a white noise, or Koike (2016) for a noise satisfying (N) and in the presence of jumps, under (K-r) for some $r < 1$.

As for the spot volatility in the presence of noise, it can basically be estimated using the local method of Sect. 4.1 after kernel de-noising, or again the Fourier method. The rates now become, at the best, $1/\Delta_n^{1/8}$.

6 When there is microstructure noise in the discontinuous case

The methods expounded in the previous section do not apply as such when X has jumps. The QMLE and MSRV probably cannot be modified in order to accommodate jumps, whereas the flat-top kernel and pre-averaging methods might perhaps be modified when X jumps, by an appropriate truncation, when jumps have finite activity and perhaps also under (H-r) for $r < 1$, but these extensions have not been explicitly done so far.

Hence, up to our knowledge, the only method devised for a jumping process X is the following one, from Jacod and Todorov (2018). This method combines pre-averaging and a method close to the one in Sect. 3.3. We assume either (K-r) with some $r < 1$ or (S) for X, (N) for the noise, and the following for the sampling scheme:

Assumption (A') We have (A) and (10), with for each n the variables $(\Phi_i^n : i \geq 1)$ being independent of $\mathcal{F}^X$ and mutually independent (not necessarily i.i.d.), with mean 1 and all moments bounded uniformly in i, n; moreover the process λ_t is càdlàg, locally bounded as well as $1/\lambda_t$, and up to localization satisfies (17).

We need a kernel f satisfying (35), two sequences k_n and ν_n of integers and a sequence u_n of positive reals, subject to the following for some $\varepsilon \in (0, 1/24)$:

$$k_n \asymp \frac{1}{\sqrt{\Delta_n}}, \qquad \Delta_n^{-1/6-\varepsilon} \leq \nu_n \leq \Delta_n^{1/4+\varepsilon}, \qquad u_n \asymp \frac{\log(1/\Delta_n)}{\Delta_n^{1/4}}. \tag{42}$$

We use the pre-averaged variables $\overline{Y}_i^n$ of (37) computed on the basis of f and the sequence k_n. We "symmetrize" the pre-averaged returns by taking the differences $\overline{Y}_i^n - \overline{Y}_{i+k_n}^n$, in order to eliminate the drift and the possible asymmetry of the jumps. We consider the logarithm of the real part of the empirical characteristic function on windows of length $2k_n\nu_n$, evaluated at yu_n for $y > 0$:

$$V(y)_i^n = -\log\left(U(y)_i^n \bigvee \frac{1}{\nu_n}\right), \quad U(y)_i^n = \frac{1}{\nu_n}\sum_{j=0}^{\nu_n-1} \cos\left(yu_n(\overline{Y}_{i+2jk_n}^n - \overline{Y}_{i+(2j+1)k_n}^n)\right).$$

For de-biasing purpose, instead of $\widehat{Y}_i^n$ we use the following

$$\widetilde{Y}_i^n = \frac{1}{\nu_n k_n}\sum_{j=1}^{\nu_n k_n}(\Delta_{i+j}^n Y)^2.$$

We aggregate those local quantities to get, with the function $g(x,y)=\frac{1}{2}\big(e^{2x-y}+e^{2x}-2\big)$ for another de-biasing and $\overline{f}_n$, $\overline{f}'_n$ as in (36):

$$\mathbb{L}(y)^n_t=\frac{2v_n}{u_n^2\,\overline{f}_n}\sum_{i=0}^{[N^n_t/(v_nk_n)]-1}\Big(V(y)^n_{iv_nk_n}-\frac{1}{2v_n}\,g\big(V(y)^n_{iv_nk_n},V(2y)^n_{iv_nk_n}\big)-\frac{\overline{f}'_n}{2k_n}\,y^2u_n^2\,\widetilde{Y}^n_{ik_nv_n}\Big).$$

Exactly as in Sect. 3.3, we have two different results.

Theorem 17 *Assume* (*K-r*) *for some* $r<1$, (*A'*) *and* (*N*), *and also* (42). *Then for any* $t>0$ *we have* $\frac{1}{\Delta_n^{1/4}}\big(\mathbb{L}(1)^n_t-C_t\big)\xrightarrow{\mathcal{L}-s}MN(0,Q^{\mathrm{cf}}_t)$, *where*

$$Q^{\mathrm{cf}}_t=4\int_0^t\left(c_s\lambda_s+\frac{\psi'(f)}{\psi(f)}\,\gamma_s\right)^2\frac{1}{\lambda_s}\,\mathrm{d}s. \tag{43}$$

Besides the presence of the process λ_s, due to the irregular sampling assumption (in the regular case we have $\lambda_t\equiv 1$), the above asymptotic variance has a rather surprising structure, quite different from the optimal asymptotic variance of (28). In the Black–Scholes, white noise and regular sampling case, and with $x=c/\gamma$, the ratio of the variance (43) and the optimal variance is $(x+12)^2/(2x^{3/2})$. The minimum of this ratio is 2.25, reached for $x=12$, and the ratio goes to ∞ when either $x\to\infty$ or $x\to 0$. So when $c=12\gamma$ we have a rather good efficiency (in the symmetric case with no drift, we could avoid differencing the definition of $U(y)^n_i$, which would give us the ratio 1.15, not very different from the optimum $\alpha'_0(f)\approx 1.07$ for the triangular kernel). But, the efficiency drops a lot when c is far from 12γ. However, we fortunately preserve the optimal rate $1/\Delta_n^{1/4}$.

To make this result feasible, we need consistent estimators $\widetilde{Q}^{\mathrm{cf},n}_t$ for Q^{cf}_t. We use a sequence of truncation levels v_n satisfying $v_n\asymp\Delta_n^{1/3}$ and set, with the notation (37):

$$\widetilde{Q}^{\mathrm{cf},n}_t=\frac{1}{k_n^2\Delta_n\,\psi(f)^2}\sum_{i=0}^{N^n_t-k_n+2}\Bigg(\frac{4}{3}\,|\overline{Y}^n_i|^4+4\left(\frac{1}{k_n^2\Delta_n}-1\right)|\overline{Y}^n_i|^2\,\widehat{Y}^n_i$$
$$+\left(\frac{1}{k_n^2\Delta_n}-1\right)^2(\widehat{Y}^n_i)^2\Bigg)\,1_{\{|\overline{Y}^n_i|+\widehat{Y}^n_i\le v_n\}},$$

and we have

$$\frac{1}{\Delta_n^{1/4}\sqrt{\widetilde{Q}^{\mathrm{cf},n}_t}}\big(\mathbb{L}(1)^n_t-C_t\big)\xrightarrow{\mathcal{L}-s}\mathcal{N}(0,1)\quad\text{in restriction to the set }\{C_t>0\}. \tag{44}$$

Observe that we still have a feasibility problem here, in general, because Δ_n is not a quantity that is observed under (A'), and so far this problem has not been solved, except in one case: namely when $\lambda_t\equiv\lambda$ is a constant (for example for regular or Poisson sampling). In this case, up to a change of scale it is no restriction to assume that $\lambda_t\equiv 1$, and although Δ_n is still not known it can be estimated by $\widehat{\Delta}_n = t/N_t^n$. Then what precedes holds with Δ_n substituted with $\widehat{\Delta}_n$, and (44) really becomes feasible.

When we only have (S) with some $\beta \geq 1$, we need the same additional de-biasing as in (21), namely

$$\widehat{C}''(y)_t^n = \mathfrak{L}(1)_t^n - \frac{\left(\mathbb{L}(y)_t^n - y^2\mathbb{L}(1)_t^n\right)^2}{\mathbb{L}(y^2)_t^n - 2y^2\mathbb{L}(y)_t^n + y^4\mathbb{L}(1)_t^n},$$

and we have basically the same result:

Theorem 18 *Assume* (*S*), (*A'*) *and* (*N*) *and* (42). *Then for any* $t > 0$ *we have* $\frac{1}{\Delta_n^{1/4}}\left(\widehat{C}''(y)_t^n - C_t\right) \xrightarrow{\mathcal{L}-s} MN(0, Q_t^{\mathrm{cf}})$, *where* Q_t^{cf} *is given by* (43).

We still have (44), except that in the definition of $\widetilde{Q}_t^{\mathrm{cf},n}$ we have to choose the truncation levels v_n such that $v_n \asymp \Delta_n^{\varpi}$ for some $\varpi \in \left(\frac{1}{4-\beta}, \frac{1}{2}\right)$ with β as in (S). Hence when β is unknown we must at least know that it does not exceed some bound $\beta_0 < 2$.

References

Ait-Sahalia, Y., Jacod, J.: High-Frequency Financial Econometrics. Princeton University Press, Princeton (2014)

Ait-Sahalia, Y., Mykland, P.: The effects of random and discrete sampling when estimating continuous-time diffusions. Econometrica **71**, 483–549 (2003)

Ait-Sahalia, Y., Mykland, P., Zhang, L.: How often to sample a continuous-time process in the present of market microstructure noise. Rev. Financ. Stud. **18**, 351–416 (2003)

Andersen, T., Dobrev, D., Schaumburg, E.: Jump-robust volatility estimation using nearest neighbor truncation. J. Econ. **169**, 75–93 (2012)

Barndorff-Nielsen, O.E., Shephard, N.: Power and bipower variation with stochastic volatility and jumps. J. Financ. Econ. **2**, 1–48 (2004)

Barndorff-Nielsen, O.E., Shephard, N., Winkel, M.: Limit theorems for multipower variation in the presence of jumps. Stoch. Process. Appl. **116**, 796–806 (2006)

Barndorff-Nielsen, O.E., Hansen, P.R., 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)

Bibinger, M., Hautsch, N., Malec, P., Reiss, M.: Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency. Ann. Stat. **42**, 1312–1346 (2014)

Clément, E., Delattre, S., Gloter, A.: An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. Stoch. Process. Appl. **123**, 2500–2521 (2013)

Clément, E., Gloter, A.: Limit theorems in the Fourier transform method for the estimation of multivariate volatility. Stoch. Process. Appl. **121**, 1097–1124 (2013)

Cuchiero, C., Teichmann, J.: Fourier transform methods for pathwise covariance estimation in the presence of jumps. Stoch. Process. Appl. **125**, 116–160 (2015)

Fan, J., Wang, Y.: Spot volatility estimation for high-frequency data. Stat. Interface **1**, 279–288 (2008)

Hayashi, T., Jacod, J., Yoshida, N.: Irregular sampling and central limit theorems for power variations: the continuous case. Ann. Inst. H. Poincaré **47**, 1197–1218 (2011)

Hayashi, T., Yoshida, N.: Asymptotic normality for a covariance estimator for nonsynchronously observed diffusions processes. Ann. Inst. Stat. Math. **60**, 367–406 (2008)

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)

Jacod, J., Li, Y., Zheng, X.: Estimating the integrated volatility with tick observations. J. Econ. **208**, 80–100 (2018)

Jacod, J., Mykland, P.: Microstructure noise in the continuous case: efficiency and the adaptive pre-averaging method. Stoch. Process. Appl. **125**, 2910–2936 (2015)

Jacod, J., Protter, P.: Discretization of Processes. Springer, Berlin (2012)

Jacod, J., Reiss, M.: A remark on the rates of convergence for integrated volatility estimation in the presence of jumps. Ann. Stat. **42**, 1131–1144 (2013)

Jacod, J., Todorov, V.: Efficient estimation of integrated volatility in presence of infinite variation jumps. Ann. Stat. **42**, 1029–1069 (2014)

Jacod, J., Todorov, V.: Limit theorems for integrated local empirical characteristic exponents from noisy high-frequency data with application to volatility and jump activity estimation. Ann. Appl. Probab. **28**, 511–576 (2018)

Koike, Y.: Estimation of integrated covariances in the simultaneous presence of nonsynchronicity, microstructure noise and jumps. Econ. Theory **32**, 533–611 (2016)

Malliavin, P., Mancino, M.-E.: Fourier series methods for measurement of multivariate volatilities. Financ. Stoch. **6**, 49–61 (2002)

Malliavin, P., Mancino, M.-E.: A Fourier method for nonparametric estimation of multivariate volatilities. Ann. Stat. **37**, 1983–2010 (2011)

Mancini, C.: Disentangling the jumps of the diffusion in a geometric Brownian motion. G. dell'Ist. Ital. degli Attuari **64**, 19–47 (2001)

Mancini, C.: The speed of convergence of the threshold estimator of integrated variance. Stoch. Process. Appl. **121**, 845–855 (2011)

Mancino, M.-E., Sanfelici, S.: Multivariate volatility estimation with high frequency data using Fourier method. In: Viens, F.G., Mariani, M.C., Florescu, I. (eds.) Handbook of Modeling High-Frequency Data in Finance, pp. 243–294. Wiley, New York (2012)

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

Renyi, A.: On stable sequences of events. Sankya Ser. A **25**, 293–302 (1963)

Vetter, M.: Limit theorems for bipower variation of semimartingales. Stoch. Process. Appl. **120**, 22–38 (2010)

Xiu, D.: Quasi-maximum likelihood estimation of volatility with high-frequency data. J. Econ. **159**, 235–250 (2010)

Zhang, L., Mykland, P., Aït-Sahalia, Y.: A tale of two time scales; determining integrated volatility with noisy high-frequency data. JASA **100**, 1394–1411 (2005)

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