Efficient estimation of integrated volatility in presence of infinite variation jumps with multiple activity indices

Jean Jacod[*] and Viktor Todorov[†]

March 15, 2015

Abstract

In a recent paper ([7]), we derived a rate efficient (and in some cases variance efficient) estimator for the integrated volatility of the diffusion coefficient of a process in presence of infinite variation jumps. The estimation is based on discrete observations of the process on a fixed time interval with asymptotically shrinking equidistant observation grid. The result in [7] is derived under the assumption that the jump part of the discretely-observed process has a finite variation component plus a stochastic integral with respect to a stable-like Lévy process with index $\beta > 1$. Here we show that the procedure of [7] can be extended to accommodate the case when the jumps are a mixture of finitely many integrals with respect to stable-like Lévy processes with indices $\beta_1 > \cdots > \beta_M \geq 1$.

AMS 2000 subject classifications: 60F05, 60F17, 60G51, 60G07.

Keywords: Central Limit Theorem, Integrated Volatility, Itô Semimartingale, Jumps, Jump Activity, Quadratic Variation, Stable Process.

1 Introduction

In this paper we revisit the question of efficient, or at least rate-efficient, estimation of the integrated volatility of the diffusion coefficient of a one-dimensional process X which is observed at discrete times. The process X is an Itô semimartingale, whose continuous martingale part is a stochastic integral $\int_0^t \sigma_s \, dW_s$ with respect to a standard Brownian motion W, and the integrated volatility over a given time interval $[0, T]$ is $C_T = \int_0^t \sigma_s^2 \, ds$.

The problem of estimating the integrated volatility C_T is arguably the most studied one in the literature dealing with inference based on high-frequency observations. The volatility inference problem has been analyzed in various settings. This includes the case when X has jumps, when the process X is observed with additional microstructure noise, and when the observations take place at irregularly spaced – and possibly random – times.

In this paper, we discard microstructure and irregular spacing, and concentrate on the case when X is perfectly observed at times $i\Delta_n$ for $i = 0, 1, \ldots$ within a fixed finite time interval $[0, T]$, and when the time lag Δ_n is small, and eventually goes to 0 (the "high-frequency" setting). We put

[*]Institut de Mathématiques de Jussieu, CNRS – UMR 7586, Université Pierre et Marie Curie–P6, 4 Place Jussieu, 75252 Paris-Cedex 05; email: jean.jacod@upmc.fr.

[†]Department of Finance, Northwestern University, Evanston, IL 60208-2001, email: v-todorov@northwestern.edu.

emphasis on the case when X has jumps with a “high” degree of activity, meaning that those jumps are not locally summable, since the case when they are summable (equivalently, with finite variation) is well established already (with references given below).

In the above setting we answer the following two questions. First, knowing that in the continuous case, and also when there are summable jumps (plus a few appropriate technical assumptions), one can exhibit estimators which converge at the optimal (or, efficient) rate $1/\sqrt{\Delta_n}$, as $\Delta_n \to 0$, is it possible to obtain estimators with the same efficient rate in the non-summable jumps case? Second, if this holds, can one find estimators that are variance-efficient, in the sense that their (normalized) asymptotic variance is the same efficient variance as the one found in the continuous case?

These two questions have been studied in the paper [7], when the jump part of X can be split into a sum $Z+Z'$ as follows: Z'_t is an absolutely convergent sum $\sum_{s\leq t}\Delta Z'_s$ (here, $\Delta V_s = V_s - V_{s-}$ is the jump size at time s for any càdlàg process V), and $Z_t = \int_0^t \gamma_s\, dY_s$ for some (unknown) process γ and a stable or “stable-like” process Y with index $\beta \in (1,2)$ (we refer to [7] for the formal definition of “stable-like” processes - they include many processes used in various applications such as the tempered stable processes). Namely, in this case in [7] we have constructed estimators which converge at the rate $1/\sqrt{\Delta_n}$, and which are variance-efficient when further the process Y is symmetric, whereas if Y is not symmetric the asymptotic variance is twice the efficient variance.

Our goal in this paper is to extend this result in two directions: first we replace the assumption $Z_t = \int_0^t \gamma_s\, dY_s$ by the fact that the “spot Lévy measures” of Y behave like a stable Lévy measure with index β near the origin, with a multiplicative coefficient which is a stochastic process a_t (see below for the precise statement); this is much weaker than assuming $Z_t = \int_0^t \gamma_s\, dY_s$. Second, we replace the single process Z by a sum $\sum_{m=1}^M Z^m$, where each Z^m is as Z above with an index β_m, and $\beta_1 > \beta_2 > \cdots > \beta_m$, and the β_m are “regularly spaced”, in the sense that they lie on a grid of the form $\{2-j\rho : j=1,\ldots,[2/\rho]\}$ for some (unknown) number $\rho>0$ (the assumption is in fact even weaker than this but too technical to be stated in the introduction). Then, we exhibit estimators for C_T which converge with the efficient rate $1/\sqrt{\Delta_n}$, and with the asymptotic variance equal to twice the efficient one. Should we have $Z^m_t = \int_0^t \gamma^m_s\, dY^m_s$ with Z^m stable-like with index β_m and symmetric, it would also be possible to construct estimators achieving the efficient variance bound, but for simplicity we will not do this here.

The method used here is basically the same as in [7], and relies on estimating locally the volatility (diffusion coefficient) from the empirical characteristic function of the increments of the process over blocks of decreasing length but containing an increasing number of observations, and then summing the local volatility estimates. The last step consists in iterating the de-biasing procedure introduced in that paper.

The rest of the paper is organized as follows. In Section 2 we describe the general setting, and present a short review of the methods which have been proposed in the literature for jump-robust volatility estimation. Section 3 is devoted to recalling the estimators of [7] and the associated CLT. In Section 4 we solve the problem of estimating integrated volatility with a rate $1/\sqrt{\Delta_n}$ under our generalized assumptions. Proofs are given in Section 5.

2 Estimating integrated volatility: a review

2.1 The general setting

The underlying process X is a one-dimensional Itô semimartingale defined on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, so it can be written as

$$X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s + \int_0^t \int_{\mathbb{R}} \delta(s,z)\, (\underline{p}(ds,dz) - \underline{q}\,(ds,dx)) + \int_0^t \int_{\mathbb{R}} \delta'(s,z)\, \underline{p}(ds,dz), \quad (2.1)$$

where W is a standard Brownian motion and $\underline{p}$ is a Poisson measure on $\mathbb{R}_+ \times E$ with (deterministic) compensator $\underline{q}\,(dt,dz) = dt \otimes \lambda(dz)$. Here E is a Polish space and λ is a σ-finite measure on E. The processes b_t and σ_t are optional, the functions δ and δ' on $\Omega \times \mathbb{R}_+ \times E$ are predictable, with $|\delta| \leq 1$ and they all are such that the integrals in (2.1) make sense (this will be implied by our assumptions below).

It is perhaps more natural to write the last two terms in (2.1) as

$$(\delta\, 1_{\{|\delta|\leq 1\}}) * (\underline{p} - \underline{q}\,)_t + (\delta\, 1_{\{|\delta|>1\}}) * \underline{p}\,_t,$$

with the same function δ in both terms. However, given the structure that we will introduce on the infinite variation jumps below (and hence on δ), the formulation (2.1) offers more flexibility. For example, it allows us to add to X a term such as $\delta' * \underline{p}$ with a function δ' having no regularity in time.

We will further assume that the volatility process σ_t is itself an Itô semimartingale, which can thus be written as

$$\begin{aligned}\sigma_t = \sigma_0 + \int_0^t b^\sigma_s\, ds + \int_0^t H^\sigma_s\, dW_s + \int_0^t H'^\sigma_s\, dW'_s + \int_0^t \int_{\mathbb{R}} \delta^\sigma(s,z)\, 1_{\{|\delta^\sigma(s,z)|\leq 1\}}\, (\underline{p} - \underline{q}\,)(ds,dz) \\ + \int_0^t \int_{\mathbb{R}} \delta^\sigma(s,z)\, 1_{\{|\delta^\sigma(s,z)|> 1\}}\, \underline{p}\,(ds,dz).\end{aligned} \tag{2.2}$$

Choosing the same Poisson measure $\underline{p}$ to drive both X and σ is not a restriction, and W' in (2.2) is another Brownian motion independent of W. Note that we need both W and W' to allow for general dependence between the diffusion components of X and σ.

Our aim in this paper is to estimate the integrated volatility which is formally defined as

$$C_T = \int_0^T c_s\, ds, \quad \text{where } c_t = \sigma_t^2.$$

The estimation is done on the basis of the observations $X_{i\Delta_n}$ for $i = 0, \cdots, [T/\Delta_n]$. We assume that $\Delta_n \to 0$, and the observed increments are denoted as $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$.

In order to describe the assumptions we need for our asymptotic results, we first introduce a property relative to a generic process U. This will be applied to U being b in (2.1) or some of the coefficients appearing in (2.2). This property is as follows: there is a constant γ such that, for all $s, t \geq 0$, we have

$$\mathbb{E}(|U_{s+t} - U_s|^2 \mid \mathcal{F}_s) \;\leq\; \gamma t. \tag{2.3}$$

When U is an Itô semimartingale with locally bounded characteristics, it satisfies this property "locally" (in the sense that for a sequence τ_n of stopping times increasing to ∞ the stopped processes $(U_{\tau_n \wedge t})_{t\geq 0}$ satisfies (2.2)). It is also satisfied when the paths of U are Hölder with index 1/2.

Our first assumption, quite standard in almost all high-frequency situations, is as follows

Assumption ($\mathbf{A}_0$): *We have* (2.1) *and* (2.2)*, and there exist two numbers $r \in [0,1)$ and $r' \in (1,2)$, a sequence τ_n of stopping times increasing to infinity, two sequences J_n, J'_n of $[0,1]$-valued Borel functions on E such that $\int J_n(z)^{r'}\, \lambda(dz) < \infty$ and $\int J_n(z)\, \lambda(dz) < \infty$, and a sequence Γ_n of numbers, such that*

$$t < \tau_n \;\Rightarrow \left\{ \begin{array}{l} |b_t|,\ |\sigma_t|,\ |b^\sigma_t|,\ |H^\sigma_t|,\ |H'^\sigma_t| \leq \Gamma_n \\ |\delta(t,z)| \leq J_n(z), \qquad |\delta'(t,z)|^r \wedge 1 \leq J'_n(z), \qquad |\delta^\sigma(t,z)|^2 \wedge 1 \leq J'_n(z), \end{array}\right. \tag{2.4}$$

and the stopped processes $H^\sigma_{t\wedge\tau_n}$ and $b_{t\wedge\tau_n}$ satisfy (2.3) *with $\gamma = \Gamma_n$.*

2.2 Some known results

As mentioned in the introduction, the question of estimating C_T for a given $T > 0$, on the basis of the discrete observations within the time interval $[0,T]$, has already been thoroughly studied. We refer to [2] for many bibliographical comments, and only provide a brief sketch below.

Let us start with the continuous case, that is, $\delta = \delta' = 0$ in (2.1). The most natural estimator at stage n is the approximate quadratic variation (or, realized volatility)

$$\widehat{C}_T^n = \sum_{i=1}^{[T/\Delta_n]} (\Delta_i^n X)^2.$$

The normalized sequence $\sqrt{\Delta_n}\,(\widehat{C}_T^n - C_t)$ converges (in the sense of stable convergence in law) to a limit Z_T which is defined on an extension of the space, and which conditionally on $\mathcal{F}$ is a centered Gaussian variable with (conditional) variance $V_T = 2\int_0^T c_s^2\,ds$, as $\Delta_n \to 0$. For this, we do not need (A_0): on top of $\delta = \delta' = 0$, having V_T and $\int_0^T b_s^2\,ds$ finite-valued is enough, and we do have functional (in T) convergence, even.

Moreover, when σ_t is non-random (so, apart from the drift term, the process X has independent increments, and is Lévy when further $\sigma_t = \sigma$ is constant), the LAN property holds for the (deterministic) parameter C_T, and V_T above is the inverse of the Fisher information, so indeed the estimators $\widehat{C}_T^n$ are asymptotically efficient. The same efficiency holds, due to the LAMN property now, when σ_t has the form $\sigma_t = \sigma(t, X_t)$ for a smooth enough function σ. Efficiency also holds, due to an infinite-dimensional Hajek type convolution theorem for more general stochastic volatility models, see [5].

These considerations lead us to postulate that "efficiency" for a sequence $\widehat{C}_T'^n$ of estimators (for X either continuous or discontinuous) means that the sequence $\sqrt{\Delta_n}\,(\widehat{C}_T'^n - C_t)$ converges stably in law to a limit which is $\mathcal{F}$-conditionally centered Gaussian with variance V_T, whereas rate-efficiency means the same with again the rate $1/\sqrt{\Delta_n}$, but with a conditional variance possibly bigger than V_T.

The above described asymptotic properties of the approximate quadratic variation, $\widehat{C}_T^n$, badly fail when X has jumps since, to begin with, $\widehat{C}_T^n$ no longer converges to C_T but to $C_T + \sum_{s \le T} (\Delta X_s)^2$. However, one can replace $\widehat{C}_T^n$ by a truncated version:

$$\widehat{C}_T'^n = \sum_{i=1}^{[T/\Delta_n]} (\Delta_i^n X)^2\, 1_{\{|\Delta_i^n X| \le u_n\}}.$$

This is the *truncated realized volatility* introduced by Mancini [9, 10]. For a properly chosen sequence of truncation level u_n (going to 0), the $\widehat{C}_T'^n$ above enjoy exactly the same properties as $\widehat{C}_T^n$ does in the continuous case, so they are efficient, under the following assumption: we have (2.1) with $\delta = 0$, and b is locally bounded, and σ is càdlàg, and δ satisfies (2.4) (recall $r < 1$). We do not need (2.2), but the key properties are that the jump part of X reduces to the last term in (2.1) and that $r < 1$.

Another way to annihilate the role of the jumps is to use the *multipower variations* introduced by Barndorff-Nielsen and Shephard, see [3], [4]:

$$\widehat{C}_T''^n = \sum_{i=1}^{[T/\Delta_n]-k+1} \prod_{m=1}^{k} |\Delta_i^n X|^{2/k},$$

where $k \ge 2$ is an integer. Then, if $k \ge 3$ and under (A_0) with $\delta = 0$ and $r \le 2/k$, plus the fact that σ never vanishes, these estimators are rate efficient, although the asymptotic variance is $\alpha(k)V_T$ for some known constant $\alpha(k) > 1$ (so we loose a bit of variance-efficiency). If instead we use the truncated bipower variation:

$$\sum_{i=1}^{[T/\Delta_n]-1} |\Delta_i^n X\, \Delta_{i+1}^n X|\, 1_{\{|\Delta_i^n X| \le u_n\}}\, 1_{\{|\Delta_{i+1}^n X| \le u_n\}}$$

for a properly chosen sequence u_n, we obtain the same result, except that we can take any $r \le 1$ in (A_0) (and still $\delta = 0$, of course), see [12].

Other methods are also available, although less widely used, and they all necessitate $\delta = 0$, or equivalently that the degree of activity of jumps (see below for this notion) is at most 1.

What happens in the presence of infinite variation jumps? We can examine for example the behavior of the truncated realized volatility $\widehat{C}'^n_T$ (the same would apply to the other estimators) when we allow $\delta \neq 0$ in (2.1). Under (A$_0$), no central limit theorem holds, but we have $\Delta_n^{\varepsilon-(2-r')/2}(\widehat{C}'^n_T - C_T) \xrightarrow{\mathbb{P}} 0$, for any $\varepsilon > 0$ (and a sequence u_n depending on ε): so the rate can approach $1/\Delta_n^{(2-r')/2}$, but not quite reaches it, and this is slower than $1/\sqrt{\Delta_n}$.

There is a good reason for this restriction, because of a minimax-type result proven in [8]: for *any* sequences of estimators $\widehat{C}'^n_t$ and positive reals w_n, if the variables $w_n\,(\widehat{C}'^n_t - C_t)$ are bounded in probability, uniformly in n and also in X ranging through all Itô semimartingales satisfying (A$_0$) for some fixed sequences Γ_n and J_n, then necessarily $w_n \leq K\big(\frac{\log(1/\Delta_n)}{\Delta_n}\big)^{(2-r')/2}$ for a constant K. This minimax rate is thus slower than $1/\sqrt{\Delta_n}$ (and becomes much slower when r' approaches 2).

Note that in (2.4) we assume $r' > 1$; of course, it may happen that one can choose J_n satisfying $\int J_n(z)^r\,\lambda(dz) < \infty$ as well, but in this case we could replace the pair (δ, δ') by the pair $(0, \delta + \delta')$. When we only have $\int J_n(z)^{r'}\,\lambda(dz) < \infty$ for some $r' > 1$, the minimax result stated above seems to rule out the existence of estimators converging to C_T at the rate $1/\sqrt{\Delta_n}$, and it certainly does in a non-parametric setting.

However, in a parametric or semi-parametric setting, things could be different. For a simpler heuristic discussion, assume that X is a Lévy process. Then $b_t = b$ and $\sigma_t = \sigma$ are constants and (A$_0$) amounts to saying that the Lévy measure F satisfies $\int(|x|^{r'} \wedge 1)\,F(dx) < \infty$: since r' is arbitrarily close to 2, this is almost the most general possible Lévy measure. When F is *known*, an analysis of the Fisher information for c, or more appropriately here for the parameter $C_T = cT$, shows that there should exist estimators with rate $1/\sqrt{\Delta_n}$ and asymptotic variance $2Tc^2$ (hence, efficient in the previous sense), see [1]. Such estimators do exist in principle (although they are not explicit), because one could show in this case that the LAN property holds. The same theoretical result holds much more generally, for Lévy processes whose Lévy measures are sums of a known measure F_0, plus any number of stable Lévy measures with indices $\beta_1 > \cdots > \beta_M$, even when those are unknown parameters, and one could prove that this holds also when F_0 is unspecified, but subject to $\int(|x|^r \wedge 1)\,F_0(dx) < \infty$ for some $r < 1$. The reason is that we are then in a semi-parametric setting, with nuisance parameters b, the β_m's and also the scale and skewness parameters of the stable Lévy measures, plus a non-parametric nuisance term F_0 which plays no role at the end because $r < 1$.

These heuristic considerations motivate the additional assumptions which are stated in the next subsection, and which are far more general than the Lévy setting of the previous paragraph. Under these assumptions we will be able to construct estimators which are at least rate-efficient even though $r' > 1$ in (A$_0$).

2.3 Additional assumptions

The additional hypotheses which are needed are expressed in terms of the spot Lévy measures of the purely discontinuous martingale term in (2.1), that is of

$$\overline{X}_t = \int_0^t \int_{\mathbb{R}} \delta(s,z)\,(\underline{p}(ds,dz) - \underline{q}\,(ds,dx)). \tag{2.5}$$

These spot Lévy measures are the measures $F_t = F_{t,\omega}$ such that the predictable compensator of the jump measure of $\overline{X}$ takes the form occurring in the $dt \otimes F_{t,\omega}(dx)$ (this factorization exists because X is an Itô semimartingale). Equivalently, one may take for $F_{t,\omega}$ the restriction to $\mathbb{R}\backslash\{0\}$ of the image of the measure λ by the map $z \mapsto \delta(\omega,t,z)$. Its symmetric tail for $x > 0$ is thus

$$\overline{F}_t(x) := F_t((-\infty,-x) \cup (x,\infty)) = \lambda(\{z : |\delta(t,z)| > x\}). \tag{2.6}$$

With this notation, our main "semi-parametric" assumption is as follows:

Assumption (A): *We have (A_0), and also an integer $M \geq 0$, a finite family $2 > \beta_1 > \cdots \beta_M > 0$ of numbers, and M nonnegative predictable càdlàg or càglàd processes $a_t^1, \ldots, a_t^m$ such that, with τ_n and Γ_n as in (A_0), the stopped processes $(a^m_{t\wedge\tau_n})^{1/\beta_m}$ and $\delta(t \wedge \tau_n, z)/J_n(z)$ for all z satisfy* (2.3) *with $\gamma = \Gamma_n$, and moreover*

$$t < \tau_n \ \Rightarrow \Big|\overline{F}_t(x) - \sum_{m=1}^M \frac{a_t^m}{x^{\beta_m}}\Big| \leq \frac{\Gamma_n}{x^r}. \tag{2.7}$$

Note that the number r' occurring in (A_0) should be bigger than β_1. When $\beta_1 < 1$ (2.7) implies for each $r'' \in (\beta_1 \vee r, 1)$ the existence of λ-integrable functions J''_n such that at $|\delta(t,z)|^{r''} \leq J''_n(z)$ when $t < \tau_n$, so in this case we could use the pair $(0, \delta + \delta')$ instead of (δ, δ') and replace r by r'': we would be on known grounds with well-established efficient estimators for C_T, as explained in the previous subsection. The same comment *a fortiori* applies when $M = 0$. So what follows has some interest only when $\beta_1 \geq 1$ and $M \geq 1$, but we use the assumption (A) as stated above for convenience in the proof. Note that, when $M = 1$, (A) is comparable, although significantly weaker, than the assumption made in [7].

(A) implies that the spot Blumenthal-Getoor index of $\overline{X}$ (BG index, for short, also called "degree of jump activity") at time t is β_m if $a_t^1 = \cdots = a_t^{m-1} = 0 < a_t^m$, and analogously the global BG index of X on the time interval $[0,T]$ is β_m if $A_T^1 = \cdots = A_T^{m-1} = 0 < A_T^m$, where we have set

$$A_T^m = \int_0^T a_s^m \, ds, \tag{2.8}$$

and the same holds for the original process X as well, upon substituting β_m with $r \vee \beta_m$. One can view the variable A_T^m as the "integrated intensity" of the jumps of $\overline{X}$ with activity degree β_m.

Remark 1 *Since we split the jumps of X into two separate components δ and δ', in order to assume some kind of regularity in time for the component δ only, it seems natural to use in (A) the spot Lévy measures of $\overline{X}$. However, since the same number r occurs in* (2.7) *and also in in* (2.4) *to control δ', it is easy to see that* (2.7) *holds for the spot Lévy measures of $\overline{X}$, if and only if it holds for the spot Lévy measures of the process X itself (those are the images of λ by the map $z \mapsto \delta(t,z) + \delta'(tz)$.*

Our representation of the jumps in X in (2.1) as integrals with respect to Poisson random measures is rather general. It accommodates specifications via Lévy-driven SDEs which are often used in applications. Indeed, any process of the form

$$X_t = X_0 + \int_0^t b_s \, ds + \int_0^t \sigma_s \, dW_s + \sum_{m=1}^M \int_0^t \gamma_{s-}^m \, dY_s^m + \delta' * \underline{p}_t \tag{2.9}$$

with W and $\underline{p}$ as in (2.1), Y^m being Lévy processes without Gaussian part and having arbitrary dependencies with $\underline{p}$ and with one another, and γ_t^m being a càdlàg adapted process, can be represented as in (2.1).

In the setting of (2.9), Assumption (A) concerning the jumps of X is satisfied as soon as the stopped processes $\gamma^m_{t\wedge\tau_n}$ satisfy (2.3) and the (non random and non time-dependent) Lévy measure F^m of Y^m satisfy $|\overline{F}^m(x) - 1/x^{\beta_m}| \leq K/x^r$ for some constant K. Note that here one has an implicit (innocuous) standardization of Z^m, and the connection with (2.7) is that $a_t^m = |\gamma_{t-}^m|^{\beta_m}$. This explains why we introduce a regularity assumption on $(a_t^m)^{1/\beta_m}$ in (A), rather than a condition on a_t itself.

If we replace $\int_0^t \gamma_{s-}^m \, dY_s^m$ by $\int_0^t \gamma_{s-}^{m,+} \, dY_s^{m,+} + \int_0^t \gamma_{s-}^{m,-} \, dY_s^{m,-}$ in (2.9) with Lévy processes $Y^{m,\pm}$ having only positive jumps, the same comments above apply and therefore the setting of [7], which uses the above spectrally positive Lévy processes, is a special case of the present one.

It turns out that we will need some more structure on the values of the successive BG indices β_m, as given by the next assumption:

Assumption (B): The numbers $2-\beta_m$ all belong to the set $\{j\rho:\ j=1,2,\cdots\}$ for some unknown constant $\rho\in(0,1)$ (so necessarily $M\leq[2/\rho]$). $\Box$

We will heavily use the following integral for β equal to one of the β_m's (it is convergent for all $\beta>0$, but absolutely convergent when $\beta>1$ only):

$$\beta>0\ \mapsto\ \chi(\beta)=\int_0^\infty\frac{\sin y}{y^\beta}\,dy. \tag{2.10}$$

Finally, the reader should be aware that, in the context of general Itô semimartingales, it is nevertheless restrictive for at least two distinct reasons:

1. (2.7) stipulates an expansion of the tail function $\overline{F}_t(x)$ near 0, and in particular that $\overline{F}_t(x)\sim a_t^1/x^{\beta_1}$. This is *not the case* for a typical Lévy measure: when F is a Lévy measure, there is a unique β ($=\beta_1$ here) such that $0<\limsup_{x\to0}x^\beta\overline{F}(x)<\infty$, but $\liminf_{x\to0}x^p\overline{F}(x)$ can vanish for $p=\beta$, and even for all $p\in(0,\beta]$. Dealing with Lévy measures F_t having such an erratic behavior near 0 seems hopeless with the method developed in [7] or below.

   So, even in the Lévy case, (2.7) is quite restrictive. Note, however, that the expansion $\sum_{m=1}^M a_t^m/x^{\beta_m}$ could be replaced by $\sum_{m=1}^M a_t^m\,L(x)/x^{\beta_m}$, where L is a slowly varying function (unspecified, but the same for all (ω,t), and also all m). This would necessitate rather obvious changes in the proofs in [7] and here.

2. In (2.7) a_t^m may be time-dependent and random. This is not the case of the β_m's, which *should be constants*. A more natural assumption would be to assume (2.7) with $\beta_m=\beta_t^m(\omega)$ being time varying and random. Unfortunately, even under strong regularity assumptions on the maps $t\mapsto\beta_t^m$, the method used here does not work at all in such a generalized setting. However, upon taking M large and ρ small in (B), we can obtain results for a reasonable (?) approximation of random and times-varying BG indices. This is perhaps a first step toward a more general theory.

3 Initial Estimators

In this section we recall the characteristic function based method for efficient estimation of integrated volatility that we proposed in [7], and we further derive an infeasible CLT for the volatility estimators in presence of infinite variation jumps with multiple BG indices. We will use the developed limit theory in the next section to propose feasible and rate efficient inference for the integrated volatility in the general setting of assumptions (A) and (B).

The volatility estimator of [7] is built from the real part of the "local" empirical characteristic functions of increments, taken at point $u_n/\sqrt{\Delta_n}$ for some sequence $u_n>0$ going to 0. Here, "local" means that the empirical characteristic function is computed on asymptotically shrinking windows of time length v_n, where $v_n=2k_n\Delta_n$ and $k_n\geq1$ is a suitable sequence of integers going to infinity, to be specified later. In particular, for each $u>0$ we set

$$L(u)_j^n=\frac{1}{k_n}\sum_{l=0}^{k_n-1}\cos(u(\Delta_{2jk_n+1+2l}^nX-\Delta_{2jk_n+2+2l}^nX)/\sqrt{\Delta_n}\,). \tag{3.1}$$

The reason for forming $L(u)_j^n$ on the basis of the first differences of the increments is to "symmetrize" the jump measure around zero as the difference of the increments of a Lévy process is equal in distribution to an increment of a Lévy process with symmetric Lévy measure. Of course, if the jumps in X are "essentially symmetric" (see [7] for formal statements), one does not need to difference the increments in the construction of $L(u)^n_j$. Note that $L(u)^n_j$ is a local version of the realized Laplace transform of volatility studied by [11].

Under (A) we have approximately

$$\mathbb{E}_{jv_n}(L(u)^n_j) \approx \exp\left(-u^2 c_{jv_n} - 2\sum_{m=1}^{M} \chi(\beta_m) u_n^{\beta_m} \Delta_n^{1-\beta_m/2} a^m_{jv_n}\right), \tag{3.2}$$

with formal results given in the proof. Since we are interested in the integrated volatility, we therefore form

$$\widehat{c}(u)^n_j = -\tfrac{1}{u^2} \log\Big(L(u)^n_j \bigvee \tfrac{1}{\log(1/\Delta_n)}\Big), \tag{3.3}$$

which satisfy $0 \le \widehat{c}(u)^n_j \le \frac{\log k_n}{u^2}$ since $L(u)^n_j$ are not bigger than 1. $\widehat{c}(u)^n_j$ serves as a local estimator of the volatility on the block, more precisely of the average of c_t over the interval $(jv_n, (j+1)v_n]$. The associated estimators for integrated volatility are then:

$$\widehat{C}(u)^n_t = v_n \sum_{j=0}^{[t/v_n]-1} \Big(\widehat{c}(u)^n_j - \frac{1}{u^2 k_n}\big(\sinh(u^2 \widehat{c}(u)^n_j)\big)^2\Big), \tag{3.4}$$

where $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$ is the hyperbolic sine. The presence of the term involving the function sinh in the volatility estimator is to eliminate the effect of an asymptotic bias which arises from the nonlinear transformation of the local empirical characteristic function on the blocks.

Under appropriate assumptions on the sequence u_n we will see that $\widehat{C}(u_n)_T$ converges to C_T, and there is an associated Central Limit Theorem with the convergence rate $1/\sqrt{\Delta_n}$. However, this CLT exhibits usually a non-negligible bias, and to account for it we consider the following normalized error processes (recall (2.10) for the function $\chi(\beta)$):

$$Z(u)^n_t = \frac{1}{\sqrt{\Delta_n}}\Big(\widehat{C}(u)^n_t - C_t - \sum_{m=1}^{M} A^m(u)^n_t\Big), \qquad A^m(u)^n_t = 2\chi(\beta_m) u_n^{\beta_m - 2} \Delta_n^{1-\beta_m/2} A^m_t. \tag{3.5}$$

The next theorem presents the CLT for $Z(u)^n_t$, and also for the differences $Z(yu)^n_t - Z(u)^n_t$ when $y > 0$. The reason for giving a CLT for these differences is that they will play a key role in the de-biasing procedure developed in the next section.

Theorem 2 *Assume (A) and let $\mathcal{Y}$ be any finite subset of $(0,\infty)$. Choose the two sequences k_n and u_n in such a way that*

$$k_n \sqrt{\Delta_n} \to 0, \qquad k_n \Delta_n^{1/2-\varepsilon} \to \infty \ \ \forall \varepsilon > 0, \qquad u_n \to 0, \qquad \frac{k_n \sqrt{\Delta_n}}{u_n^2} \to 0; \tag{3.6}$$

Then, we have the (functional) stable convergence in law:

$$\Big(Z(u_n)^n, \big(\frac{1}{u_n^2}(Z(yu_n)^n - Z(u_n)^n)\big)_{y\in\mathcal{Y}}\Big) \stackrel{\mathcal{L}-s}{\Longrightarrow} \big(Z, ((y^2-1)\overline{Z})_{y\in\mathcal{Y}}\big), \tag{3.7}$$

where the limit is defined on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}}_t)_{t\ge 0}, \widetilde{\mathbb{P}})$ of the original space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$ and can be written as

$$Z_t = 2\int_0^t c_s\, dW^{(1)}_s, \qquad \overline{Z}_t = \frac{2}{\sqrt{3}}\int_0^t c_s^2\, dW^{(2)}_s. \tag{3.8}$$

where $W^{(1)}$ and $W^{(2)}$ are two independent Brownian motions, independent of the σ-field $\mathcal{F}$.

Theorem 2 extends a corresponding result in [7] for $\widehat{C}(u)_t^n$ to the setting of jumps with multiple BG indices. We note that for the above theorem we assume only assumption (A) and we do not need assumption (B) for the relationship between the values of the multiple BG indices.

The two limiting processes Z and $\overline{Z}$ can be equivalently characterized by the fact that, defined on the extended space, they are globally $\mathcal{F}$-conditional centered Gaussian martingales, $\mathcal{F}$-conditionally independent, with $\mathcal{F}$-conditional variance

$$\widetilde{\mathbb{E}}((Z_t)^2 \mid \mathcal{F}) = 4\int_0^t c_s^2\, ds, \qquad \widetilde{\mathbb{E}}((\overline{Z}_t)^2 \mid \mathcal{F}) = \frac{4}{3}\int_0^t c_s^4\, ds. \tag{3.9}$$

We note that since $u_n \to 0$, (3.7) implies that the difference $Z(yu_n)^n - Z(u_n)^n$ for some $y \neq 1$ is of higher asymptotic order. We will make use of this fact when developing the debiasing procedure in the next section. A consequence of the above result is the following degenerate asymptotic behavior:

$$(Z(y_l u_n)^n)_{1\le l\le L} \stackrel{\mathcal{L}-s}{\Longrightarrow} (Z,\cdots,Z). \tag{3.10}$$

Finally, concerning the choices of the sequences k_n and u_n, one can set $k_n \asymp 1/\sqrt{\Delta_n}\,(\log(1/\Delta_n))^\alpha$ and $u_n \asymp 1/(\log(1/\Delta_n))^{\alpha'}$, for any reals α, α' such that $0 < \alpha' < \frac{\alpha}{2}$.

4 Efficient Estimators

When $M = 0$ or $\beta_1 < 1$ the estimators $\widehat{C}(u_n)_t^n$ converge to C_t at rate $1/\sqrt{\Delta_n}$, and there is no asymptotic bias. When $M \geq 1$ and $\beta_1 \geq 1$ we still have the convergence, but the rate now is arbitrarily close to, although slower than, $1/\Delta_n^{(2-\beta_1)/2}$ (when expressed as a power of Δ_n): so the estimators almost achieve the minimax bound found in [8], and the rate $1/\sqrt{\Delta_n}$ is not achieved. In view of Theorem 2, the reason for this slow rate of convergence is the presence of biases coming from the infinite jump variation part of the process. One can therefore estimate the biases and perform the following bias correction which restores rate efficient estimation of integrated volatility. We fix the time horizon $T > 0$, pick some $\zeta > 1$, and then set

$$\widehat{C}(u,\zeta)_T^n = \widehat{C}(u)_T^n - \frac{(\widehat{C}(\zeta u)_T^n - \widehat{C}(u)_T^n)^2}{\widehat{C}(\zeta^2 u)_T^n - 2\widehat{C}(\zeta u)_T^n + \widehat{C}(u)_T^n}. \tag{4.1}$$

In [7], we proved the following result for the bias-corrected estimator $\widehat{C}(u,\zeta)_T^n$. However, as already mentioned, the assumption in that paper is stronger than (A), even when $M = 1$, and it is also assumed that $\beta_1 > 1$, so we will have to revisit the proof this theorem:

Theorem 3 *Assume (A) with $M = 1$ and also that $C_T > 0$ a.s. Assume that* (3.6) *holds for k_n, u_n. Then the variables $\frac{1}{\sqrt{\Delta_n}}\,(\widehat{C}(u_n,\zeta)_T^n - C_T)$ converge stably in law to the variable Z_T.*

Notice the additional assumption $C_T > 0$ a.s. in Theorem 3. This assumption is necessary in order for the denominators in (4.1) to be "non-degenerate" in an appropriate sense. These denominators could indeed go to 0 with an uncontrolled speed on the set $\{C_T = 0\}$ on which the Brownian part of X is not active between times 0 and T, so indeed if $\mathbb{P}(C_T > 0) < 1$ the above result holds in restriction to the set $\{C_T > 0\}$ only.

Of course, when $\beta_1 < 1$ there is no need to de-bias the initial estimators $\widehat{C}(u_n)_T^n$, but *a priori* we do not know whether $\beta_1 < 1$ or not. As a matter of fact, we do not know either whether $M = 1$ or not, and when $M \geq 2$ the above de-biasing is in general not sufficient to remove all the bias due to the infinite variation jumps which is of higher order than $1/\sqrt{\Delta_n}$. Under the additional assumption (B), however, this can be done by "iterating" the previous de-biasing method. The generalized de-biasing procedure goes as follows, for any given integer N:

Step 1 - initialization: Choose a real $\zeta > 1$ and an integer $k \geq 1$, and put $\widehat{C}(u,\zeta,0)_T^n = \widehat{C}(u)_t^n$.

Step 2 - iteration: Assuming $\widehat{C}_n(u,\zeta,j-1)$ known for some integer j between 1 and k, define $\widehat{C}_n(u,\zeta,j)$ as

$$\widehat{C}(u,\zeta,j)^n_T = \widehat{C}_n(u,\theta,j-1)^n_T + \frac{(\widehat{C}_n(\zeta u,\theta,j-1)^n_T - \widehat{C}(u,\zeta,j-1)^n_T)^2}{\widehat{C}(\zeta^2 u,\zeta,j-1)^n_T - 2\widehat{C}(\theta u,\zeta,j-1)^n_T + \widehat{C}(u,\theta,j-1)^n_T} + u^2\sqrt{\Delta_n}. \tag{4.2}$$

Step 3: The final estimator is set to be $\widehat{C}(u_n,\zeta,N)^n_T$.

The asymptotic result for $\widehat{C}(u_n,\zeta,N)^n_T$ is given in the following theorem.

Theorem 4 *Assume (A) and (B) with $\rho \geq \rho_0$ for some $\rho_0 \in (0,1)$, and also that $C_T > 0$ a.s. Let N be the biggest integer such that $N\rho_0 \leq 1$. Take any $\zeta > 1$ and choose the sequences k_n and u_n satisfying* (3.6)*. Then the variables $\frac{1}{\sqrt{\Delta_n}}(\widehat{C}(u_n,\zeta,N)^n_T - C_T)$ converge stably in law to the variable Z_T.*

In practice, of course, we don't know ρ, but we can decide the number of iterations we are ready to undertake. This amounts to choosing *a priori* the value of N, and then assume that (B) holds with some $\rho \geq 1/N$. If the "real" ρ is bigger than $1/N'$ for some integer $N' < N$, then we have iterated the procedure too many times than is actually necessary. This does not harm from an asymptotic viewpoint.

However, in finite samples de-biasing can make the estimator very unstable. Therefore, in practice it is desirable to have a data-driven adaptive choice of N which does not perform more de-biasing than is actually needed. Developing such a method can be based on the difference $\widehat{C}_n(u,\theta,j)^n_T - \widehat{C}(\zeta u,\zeta,j)^n_T$ and its asymptotic distribution when $j\rho_0 \leq 1$ which given our previous results is asymptotically mixed-normal. In an adaptive method for de-biasing, one will proceed debiasing until the difference $\widehat{C}_n(u,\theta,j)^n_T - \widehat{C}(\zeta u,\zeta,j)^n_T$ becomes smaller than some high quantile of its limit distribution. If j is too small, then the bias term will dominate the limiting behavior of $\widehat{C}_n(u,\theta,j)^n_T - \widehat{C}(\zeta u,\zeta,j)^n_T$ and hence will eventually exceed the above-mentioned high quantile of the limit distribution. We leave the formal analysis of such adaptive method for future work.

Remark 5 *The estimators $\widehat{C}(u_n,\zeta,N)^n_T$ are rate-efficient, but not variance-efficient, since their asymptotic variance is twice the (supposedly) optimal one $2\int_0^T c_s^2\,ds$, see subsection 2.2. This comes of course from the fact that in* (3.1) *we take the differences of two successive increments. So, we really use about $[T/2\Delta_n]$ variables (the afore-mentioned differences) instead of all the $[T/\Delta_n]$ increments.*

As explained in [7], it is possible to use

$$L(u)^n_j = \frac{1}{k_n}\sum_{l=0}^{k_n-1} \cos(u\Delta^n_{jk_n+1+l}X/\sqrt{\Delta_n}\,) \tag{4.3}$$

instead of (3.1)*, and accordingly modify the de-biasing terms and the normalization in* (3.3)*, in the case the jumps are "essentially symmetric" about* 0*: this means that we have the same expansion, with the same processes $a^m_t/2$, for the left and right tails $\overline{F}^-_t(x) = F_t((-\infty,-x))$ and $\overline{F}^+_t(x) = F_t((x,\infty))$. This would result in getting estimators achieving the optimal asymptotic variance.*

On the other hand, if we use (4.3) *instead of* (3.1) *when this symmetry assumption fails, we still have Theorem 2, but with a centering term in* (3.5) *having a much more complicated structure. When there is a single index ($M = 1$) it is possible to do an iteration procedure resulting in a theorem analogous to the previous one, but now with the optimal conditional variance. Otherwise, since the normalizing factors $u_n^{\beta_m-2}\Delta_n^{1-\beta_m/2}$ in* (3.5) *and those occurring in the new additional centering terms are not commensurable, a simple iteration procedure as the one described above is not going to work.*

5 Proofs

A standard localization procedure shows that it is enough to prove all results of this paper when (A) is substituted with

Assumptions (SA): *We have (A) with $\tau_1 \equiv \infty$, we write $J = J_1$ and $J' = J'_1$, and moreover we have $|\delta'(.,z)|^r \leq \Gamma J(z)$ and $|\delta^\sigma(.z)|^2 \leq \Gamma J(z)$ for some constant Γ (implying in particular that δ' and δ^σ are bounded).*

Below, this strengthened assumption is in force. Up to modifying b^σ without altering its boundedness, one may incorporate the last integral in (2.2) into the previous compensated sum of jumps term, and also use Itô's formula, to get for suitable processes and function b^c, H^c, H'^c, δ^c:

$$\begin{array}{l} \sigma_t = \sigma_0 + \int_0^t b_s^\sigma\, ds + \int_0^t H_s^\sigma\, dW_s + \int_0^t H_s'^\sigma\, dW'_s + \int_0^t \int_E \delta^\sigma(s,z)\,(\underline{p} - \underline{q}\,)(ds,dz) \\ c_t = c_0 + \int_0^t b_s^c\, ds + \int_0^t H_s^c\, dW_s + \int_0^t H_s'^c\, dW'_s + \int_0^t \int_E \delta^c(s,z)\,(\underline{p} - \underline{q}\,)(ds,dz). \end{array} \tag{5.1}$$

Then (SA) implies, for some constant $\Gamma \geq 1$,

$$\begin{array}{l} |b_t|,\, |\sigma_t|,\, |c_t|,\, |b_t^\sigma|,\, |b_t^c|,\, |H_t^\sigma|,\, |H_t'^\sigma|,\, |H_t^c|,\, |H_t'^c|,\, a_t \leq \Gamma \\ |\delta(t,z)|^{r'} \leq J(z), \qquad |\delta'(t,z)|^r,\, |\delta^\sigma(t,z)|^2,\, |\delta^c(t,z)|^2 \leq \Gamma J'(z) \\ V_t = \widetilde{b}_t,\ H_t^\sigma,\ H_t^c,\ \frac{\delta(t,z)}{J(z)} \qquad \Rightarrow \quad \mathbb{E}(|V_{s+t} - V_s|^2 \mid \mathcal{F}_s) \leq \Gamma t \\ V_t = \sigma_t,\ c_t \quad \Rightarrow \quad |\mathbb{E}(V_{s+t} - V_s \mid \mathcal{F}_s)| \leq \Gamma t,\ \ \mathbb{E}(|V_{s+t} - V_s|^2 \mid \mathcal{F}_s) \leq \Gamma t \\ \overline{F}_t(x) \leq \frac{K}{x^{\beta_1}}, \qquad \left|\overline{F}_t(x) - \sum_{m=1}^M \frac{a_t^m}{x^{\beta_m}}\right| \leq \frac{\Gamma}{x^r} \\ \mathbb{E}\big(|a_{s+t}^m - a_s^m|^{2/(\beta_m \wedge 1)} \mid \mathcal{F}_s\big) \leq \Gamma t \end{array} \tag{5.2}$$

(for the last line we use that, if $x, y \in [-\alpha, \alpha]$, we have $\big||x|^\beta - |y|^\beta\big| \leq |x-y|^\beta$ when $0 < \beta \leq 1$, whereas $\big||x|^\beta - |y|^\beta\big| \leq \beta\,\alpha^{\beta-1}|x-y|$ when $\beta > 1$).

Below, y, y' implicitly are in the fixed finite subset $\mathcal{Y}$ of $(0,\infty)$, which is supposed to contain 1. We denote a generic constant, changing from line to line, as K, and it possibly depends on r, M, β_m, Γ, and sometimes on some extra parameter q such as a power or on the set $\mathcal{Y}$, but never on n and the various indices $i, j, \ldots$ or variables $u, y, \ldots$ which may occur. Analogously, ϕ_n stands for a generic sequence decreasing to 0.

We will also use the following trick: if $\beta_1 \leq 1$, we add a fictitious index $\beta_0 \in (1,2)$, with the associated process a_t^0 vanishing identically. The conditions about $\overline{F}_t$ in (5.2) are satisfied as well if we replace β_1 by β_0 for the first one, and the sum $\sum_{m=1}^M$ by $\sum_{m=0}^M$ for the second one, whereas the processes $A^0(u)$ in (3.5) also vanish identically, so that the sum $\sum_{m=M_0}^M A^m(u)_t$ is the same with $M_0 = 1$ and $M_0 = 0$. Note also that β_0 can be chosen arbitrarily in $(1,2)$. Hence, we can and will assume, without restriction, that

$$\frac{3}{2} < \beta_1 < 2. \tag{5.3}$$

5.1 Estimates

The two sequences k_n, u_n, satisfying (3.6), are fixed. As in [7] the following processes play a key role:

$$\begin{array}{c} U(u)_t = e^{-u^2 c_t}, \qquad \overline{U}(m;u)_t^n = e^{-2\Delta_n^{1-\beta_m/2}\chi(\beta_m) u^{\beta_m}\, a_t^m} \\ \mathcal{U}(u)_t^n = U(u)_t \prod_{m=1}^M \overline{U}(m;u)_t^n. \end{array} \tag{5.4}$$

Since $0 \leq c_t \leq \Gamma$ and $0 \leq a_t^m \leq \Gamma$ we see that, upon increasing Γ if necessary, we have for all $u \in (0,1]$:

$$\frac{1}{\Gamma} \leq U(u)_t \leq 1, \quad \frac{1}{\Gamma} \leq \overline{U}(m;u)_t^n \leq 1, \quad \frac{1}{\Gamma} \leq \mathcal{U}(u)_t^n \leq 1. \tag{5.5}$$

According to (6.12) and (6.13) of [7], plus a trivial change due to the fact that our assumptions on a_t^m yield that $|\mathbb{E}(a_{t+s}^m - a_s \mid \mathcal{F}_s)|$ is smaller than $Kt^{(\beta_m\wedge 1)/2}$ instead of $Kt^{\beta_m/2}$, we have for all $q \geq 2$ and $u \in (0,1]$:

$$\begin{array}{l} |\mathbb{E}(U(u)_{t+s} - U(u)_t \mid \mathcal{F}_t)| \leq Ku^2 s, \qquad \mathbb{E}(|U(u)_{t+s} - U(\kappa,u)_t|^q \mid \mathcal{F}_t) \leq Ku^{2q} s \\ |\mathbb{E}(\overline{U}^m(u)_{t+s}^n - \overline{U}^m(u)_t^n \mid \mathcal{F}_t)| \leq K\Delta_n^{1-\beta_m/2} u^{\beta_m} s^{(\beta_m\wedge 1)/2} \\ \mathbb{E}(|\overline{U}^m(u)_{t+s}^n - \overline{U}^m(u)_t^n|^q \mid \mathcal{F}_t) \leq K\Delta_n^{q(1-\beta_m/2)} u^{q\beta_m} s^{(q\beta_m/2)\wedge 1} \end{array} \tag{5.6}$$

and thus

$$\begin{array}{l} |\mathbb{E}(\mathcal{U}(u)_{t+s}^n - \mathcal{U}(u)_t^n \mid \mathcal{F}_t)| \leq K\big(u^2 s + \sum\limits_{m=1}^M \Delta_n^{1-\beta_m/2} u^{\beta_m} s^{\beta_m/2}\big) \\ \mathbb{E}(|\mathcal{U}(u)_{t+s}^n - \mathcal{U}(u)_t^n|^q \mid \mathcal{F}_t) \leq K\big(u^{2q} s + \sum\limits_{m=1}^M \Delta_n^{q(1-\beta_m/2)} u^{q\beta_m} s^{1\wedge(q\beta_m/2)}\big) \\ \mathbb{E}(|\mathcal{U}(u)_{t+s}^n - \mathcal{U}(u)_t^n - (U(u)_{t+s} - U(u)_t)\prod_{m=1}^M U(m;u)_t^n|^q \mid \mathcal{F}_t) \\ \qquad\qquad \leq K \sum\limits_{m=1}^M s^{1\wedge(q\beta_m/2)}\, \Delta_n^{q(1-\beta_m/2)} u^{q\beta_m}. \end{array} \tag{5.7}$$

Next, with the notation $\psi_t^{n,i} = 1_{(i\Delta_n,(i+1)\Delta_n]}(t) - 1_{((i+1)\Delta_n,(i+2)\Delta_n]}(t)$, we set

$$\begin{array}{l} \rho_i'^n = \frac{u_n}{\sqrt{\Delta_n}}\, \sigma_{i\Delta_n}(\Delta_{i+1}^n W - \Delta_{i+2}^n W) = \frac{u_n}{\sqrt{\Delta_n}} \int_{R_+\times E} \sigma_{i\Delta_n}\psi_t^{n,i}\, dW_t \\ \rho_i''^n = \frac{u_n}{\sqrt{\Delta_n}} \int_{R_+\times E} \delta(i\Delta_n,z)\psi_t^{n,i}\,(\underline{p} - \underline{q}\,)(dt,dz) \\ \rho_i^n = \rho_i'^n + \rho_i''^n, \qquad \overline{\rho}_i^n = \frac{1}{u_n}\,(\Delta_{i+1}^n X - \Delta_{i+2}^n X), \end{array}$$

and also

$$\xi(y)_j^{w,n} = \left\{ \begin{array}{ll} \frac{1}{k_n}\sum_{l=0}^{k_n-1}\big(\cos(y\rho_{1+2jk_n+2l}^n) - \mathcal{U}(yu_n)_{2(jk_n+l)\Delta_n}^n\big) & \text{if } w=1 \\ \frac{1}{k_n}\sum_{l=0}^{k_n-1}\big(\cos(y\overline{\rho}_{1+2jk_n+2l}^n) - \cos(y\rho_{1+2jk_n+2l}^n)\big) & \text{if } w=2 \\ \frac{1}{k_n}\sum_{l=0}^{k_n-1}(\mathcal{U}(yu_n)_{\kappa(jk_n+l)\Delta_n}^n - \mathcal{U}(yu_n)_{jv_n}^n) & \text{if } w=3. \end{array}\right. \tag{5.8}$$

$$\xi(y)_j^n = \frac{1}{\mathcal{U}(yu_n)_{jv_n}^n}\sum_{w=1}^3 \xi(y)_j^{w,n}$$

$$\Omega(y)_{n,t} = \bigcap_{0\leq j<[t/v_n]}\{|\xi(y)| \leq \tfrac{1}{2}\}, \qquad \Omega_{n,t} = \bigcap_{y\in\mathcal{Y}} \Omega(y)_{n,t}.$$

Lemma 6 *For all $q \geq 2$ and $y \in \mathcal{Y}$ we have*

$$\begin{array}{l} \big|\mathbb{E}(\cos(y\overline{\rho}_i^n) - \cos(y\rho_i^n) \mid \mathcal{F}_{i\Delta_n})\big| \leq u_n^4\sqrt{\Delta_n}\,\phi_n \\ \mathbb{E}(|\cos(y\overline{\rho}_i^n) - \cos(y\rho_i^n)|^q \mid \mathcal{F}_{i\Delta_n}) \leq u_n^4\sqrt{\Delta_n}\,\phi_n. \end{array} \tag{5.9}$$

Proof. This is basically Lemma 11 of [7], with a few changes which we explain below. We have $y\overline{\rho}_i^n = \sum_{k=1}^5 \theta(k)_i^n$, where

$$\begin{array}{l} \theta(1)_i^n = y\rho_i^n, \qquad \theta(2)_i^n = \frac{yu_n}{\sqrt{\Delta_n}}\int_{R_+}(\sigma_s - \sigma_s^{n,i})\psi_s^{n,i}\,dW_s, \qquad \theta(3)_i^n = \frac{yu_n}{\sqrt{\Delta_n}}\int_{R_+}(b_s - b_{i\Delta_n})\psi_s^{n,i}\,ds \\ \theta(4)_i^n = \frac{yu_n}{\sqrt{\Delta_n}}\int_{R_+\times E}(\delta(s,z) - \delta(i\Delta_n,z))\psi_s^{n,i}\,(\underline{p} - \underline{q}\,)(ds,dz) \\ \theta(5)_i^n = \frac{yu_n}{\sqrt{\Delta_n}}\int_{R_+\times E}\delta'(s,z)\psi_s^{n,i}\,\underline{p}(ds,dz). \end{array}$$

The only difference with [7] is in the definition of the term $\theta(4)_i^n$, so the proof goes through, provided we have the following estimate:

$$\mathbb{E}(|\theta(4)_i^n| \wedge 1 \mid \mathcal{F}_{i\Delta_n}) \leq u_n^4\sqrt{\Delta_n}\,\phi_n. \tag{5.10}$$

Let $v \in (0,1]$, and recall $|\delta(s,z)| \leq J(z) \leq 1$. We have $\theta(4)_i^n = M(v)^n + N(v)^n - B(v)^n$, where

$$\begin{array}{l} M(v)^n = \frac{yu_n}{\sqrt{\Delta_n}}\int_{R_+\times\{z:J(z)\leq v\}}(\delta(s,z) - \delta(i\Delta_n,z))\psi_s^{n,i}\,(\underline{p} - \underline{q}\,)(ds,dz) \\ B(v)^n = \frac{yu_n}{\sqrt{\Delta_n}}\int_{R_+\times\{z:J(z)>v\}}(\delta(s,z) - \delta(i\Delta_n,z))\psi_s^{n,i}\,\underline{q}\,(ds,dz) \\ N(v)^n = \frac{yu_n}{\sqrt{\Delta_n}}\int_{R_+\times\{z:J(z)>v\}}(\delta(s,z) - \delta(i\Delta_n,z))\psi_s^{n,i}\,\underline{p}(ds,dz), \end{array}$$

First, upon using (5.2) and $\int_E J(z)^{2-\varepsilon}\,\lambda(dz)<\infty$ for some $\varepsilon>0$, we get

$$\begin{array}{lcl}\mathbb{E}\big((M(v)^n)^2\mid\mathcal{F}_{i\Delta_n}\big) &=& \frac{y^2u_n^2}{\Delta_n}\mathbb{E}\Big(\int_{i\Delta_n}^{(i+2)\Delta_n}ds\int_{\{z:J(z)\leq v\}}\frac{|\delta(s,z)-\delta(i\Delta_n,z)|^2}{J(z)^2}\,J(z)^2\,\lambda(dz)\mid\mathcal{F}_{i\Delta_n}\Big)\\ &\leq& Ku_n^2\Delta_n v^\varepsilon\int_E J(z)^{2-\varepsilon}\,\lambda(dz)\ \leq\ Ku_n^2\Delta_n v^\varepsilon.\end{array}$$

Second, we have

$$\begin{array}{lcl}\mathbb{E}(|B(v)^n|\mid\mathcal{F}_{i\Delta_n}) &\leq& \frac{yu_n}{\sqrt{\Delta_n}}\mathbb{E}\Big(\int_{i\Delta_n}^{(i+2)\Delta_n}ds\int_{\{z:J(z)>v\}}\frac{|\delta(s,z)-\delta(i\Delta_n,z)|}{J(z)}\,J(z)\,\lambda(dz)\mid\mathcal{F}_{i\Delta_n}\Big)\\ &\leq& K\frac{u_n\Delta_n}{v^{1-\varepsilon}}\int_E J(z)^{2-\varepsilon}\,\lambda(dz)\ \leq\ K\frac{u_n\Delta_n}{v^{1-\varepsilon}}.\end{array}$$

Third, the process $\underline{p}([0,t]\times\{z:J(z)>v\})$ is a Poisson process with parameter $\lambda(\{z:J(z)>v\})\leq K/v^{2-\varepsilon}$, hence

$$\mathbb{P}(N(v)^n\neq 0\mid\mathcal{F}_{i\Delta_n})\ \leq\ K\,\frac{\Delta_n}{v^{2-\varepsilon}}.$$

Therefore, the left hand side of (5.10) is not bigger than

$$\mathbb{E}(|M(v)^n|\mid\mathcal{F}_{i\Delta_n})+\mathbb{E}(|B(v)^n|\mid\mathcal{F}_{i\Delta_n})+\mathbb{P}(N(v)^n\neq 0\mid\mathcal{F}_{i\Delta_n})\leq K\big(v^{\varepsilon/2}\sqrt{\Delta_n}+v^{\varepsilon-2}\Delta_n\big)$$

by the previous estimates and the Cauchy-Schwarz inequality. Upon taking $v=v_n=\Delta_n^{2/(4-\varepsilon)}$ and using the consequence of (3.6) which is $\Delta_n^\eta\leq Ku_n$ for any $\eta>0$, we deduce (5.10). $\square$

Lemma 7 *For all $q\geq 2$ we have*

$$\big|\,\mathbb{E}(\cos(yu_n\rho_i^n)\mid\mathcal{F}_{i\Delta_n})-\mathcal{U}(yu_n)^n_{i\Delta_n}\big|\ \leq\ u_n^4\sqrt{\Delta_n}\,\phi_n \tag{5.11}$$

$$\Big|\,\mathbb{E}(\cos(yu_n\rho_i^n)\,\cos(y'u_n\rho_i^n)\mid\mathcal{F}_{i\Delta_n})-\frac{\mathcal{U}((y+y')u_n)^n_{i\Delta_n}+\mathcal{U}(|y-y'|u_n)^n_{i\Delta_n}}{2}\Big|\leq u_n^4\sqrt{\Delta_n}\,\phi_n \tag{5.12}$$

$$\mathbb{E}(|\cos(yu_n\rho_i^n)-\mathcal{U}(yu_n)^n_{i\Delta_n}|^q\mid\mathcal{F}_{i\Delta_n})\ \leq\ Ku_n^4. \tag{5.13}$$

Proof. In view of the definition of ρ_i^n and of the property that F_t is the image of λ by the map $z\mapsto\delta(t,z)$, we see that, with the notation $G(w)_t=\int(1-\cos(wx))\,F_t(dx)$, we have

$$\mathbb{E}(\cos(yu_n\rho_i^n)\mid\mathcal{F}_{i\Delta_n})=U(yu_n)_{i\Delta_n}\,e^{-2\Delta_nG(yu_n/\sqrt{\Delta_n})_{i\Delta_n}}. \tag{5.14}$$

The analysis of the asymptotic behavior of $G(w)_t$ as $w\to\infty$ is similar with what is done in the proof of Lemma 12 of [7]. Letting $\zeta\in(0,1]$ and using (5.2) we first observe that

$$0\leq\int_{\{x:\,|x|>\zeta\}}(1-\cos(wx))\,F_t(dx)\leq 2\overline{F}_t(\zeta)\leq\frac{K}{\zeta^{\beta_1}}.$$

Next, Fubini's theorem yields

$$\int_{\{x:\,|x|\leq\zeta\}}(1-\cos(wx))\,F_t(dx)=\int_0^{\zeta w}\overline{F}_t\big(\frac{z}{w}\big)\,\sin(z)\,dz-\int_0^{\zeta w}\overline{F}_t(\zeta)\,\sin(z)\,dz,$$

and the absolute value of the last term above is again smaller than K/ζ^{β_1} because $\big|\int_0^x\sin z\,dz\big|\leq 2$ for all x. To evaluate the first term we use (5.2) again to get

$$\Big|\int_0^{\zeta}\overline{F}_t\big(\frac{z}{w}\big)\,\sin(z)\,dz-\sum_{m=1}^M w^{\beta_m}\,\chi(\beta_m)a_t^m\Big|\leq\sum_{m=1}^M\Big|w^{\beta_m}\int_{\zeta w}^\infty\frac{\sin z}{z^{\beta_m}}\,dz\Big|+K\int_0^{\zeta w}\frac{w^r}{z^r}\,dz.$$

We have $\int_x^\infty \frac{\sin z}{z^{\beta_m}}\,dz = \frac{\cos x}{x^{\beta_m}} - \beta_m \int_x^\infty \frac{\cos z}{z^{1+\beta_m}}\,dz$ by integration by parts, yielding $\big|\int_x^\infty \frac{\sin z}{z^{\beta_m}}\,dz\big| \leq 2/x^{\beta_m}$, whereas $\int_0^{\zeta w} \frac{w^r}{z^r}\,dz = \frac{1}{1-r}\,w\zeta_n^{1-r}$ (recall $r<1$). Putting all these together yields

$$\Big|G(w)_t - \sum_{m=1}^M w^{\beta_m}\,\chi(\beta_m)a_t^m\Big| \leq K\Big(\zeta^{-\beta_1} + \sum_{m=1}^M \zeta^{-\beta_m} + w\,\zeta^{1-r}\Big) \leq K(\zeta^{-\beta_1} + w\zeta^{1-r})$$

because $\beta_m \leq \beta_1$. We apply this with $w = yu_n/\sqrt{\Delta_n}$ and $\zeta = \Delta_n^{1/2(1-r+\beta_1)}$ to get

$$\big|2\Delta_n G(yu_n/\sqrt{\Delta_n}\,)_t - \sum_{m=1}^M \Delta_n^{1-\beta_m/2}(yu_n)^{\beta_m}\chi(\beta_m)\,a_t^m\big| \leq K\Delta_n^{\frac{2-2r+\beta_1}{2-2r+2\beta_1}} \leq Ku_n^4\sqrt{\Delta_n},$$

because again $\Delta_n^\eta \leq Ku_n$ for any $\eta>0$, whereas $\frac{2-2r+\beta_1}{2-2r+2\beta_1} > \frac{1}{2}$. Since $|e^x - e^y| \leq |x-y|$ if $x,y \leq 0$, and recalling (5.4) and (5.14), we deduce (5.11).

At this stage, (5.12) and (5.13) follow from (5.11) exactly as in Lemma 12 of [7]. $\square$

Below, we use the simplifying notation

$$V(y,y')_t^n = \mathcal{U}((y+y')u_n)_t^n + \mathcal{U}(|y-y'|u_n)_t^n - 2\mathcal{U}(yu_n)_t^n\,\mathcal{U}(y'u_n)_t^n. \tag{5.15}$$

Lemma 13 of [7] is unchanged, except for (6.38) which becomes here for any $q \geq 2$:

$$\begin{array}{rcl} |\mathbb{E}(\xi(yu_n)_j^{3,n} \mid \mathcal{F}_{jv_n})| & \leq & Ku_n^4\sqrt{\Delta_n}\,\phi_n \\ \mathbb{E}(|\xi(yu_n)_j^{3,n}|^q \mid \mathcal{F}_{jv_n}) & \leq & K\big(u_n^{2q}v_n + \sum_{m=1}^M \Delta_n^{q(1-\beta_m/2)}u_n^{q\beta_m}v_n^{1\wedge(q\beta_m/2)}\big). \end{array} \tag{5.16}$$

However, when we assume (5.3), that is $\beta_1 > \frac{3}{2}$, it is clear that $\Delta_n^{q(1-\beta_m/2)}u_n^{q\beta_m}v_n^{1\wedge(q\beta_m/2)} \leq K\Delta_n^{q(1-\beta_1/2)}u_n^{q\beta_1}v_n$ for all $m \geq 1$. So, we indeed have (6.38) of [7], without change at all, and the following key lemma follows:

Lemma 8 *For all $q \geq 2$ we have for $j < [t/v_n]$:*

$$\begin{array}{rcl} |\mathbb{E}(\xi(yu_n)_j^n \mid \mathcal{F}_{jv_n})| & \leq & Ku_n^4\sqrt{\Delta_n}\,\phi_n \\ \big|\mathbb{E}(\xi(yu_n)_j^n\,\xi(y'u_n)_j^n \mid \mathcal{F}_{jv_n}) - \frac{1}{2k_n}\frac{V(y,y')_{jv_n}^n}{\mathcal{U}(yu_n)_{jv_n}^n\mathcal{U}(y'u_n)_{jv_n}^n}\big| \leq Ku_n^4\sqrt{\Delta_n}\,\phi_n \\ \mathbb{E}(|\xi(yu_n)_j^n|^q \mid \mathcal{F}_{jv_n}) & \leq & K\big(\frac{u_n^4}{k_n^{q/2}} + u_n^{2q}v_n + \Delta_n^{q(1-\beta_1/2)}u_n^{q\beta_1}v_n\big). \end{array} \tag{5.17}$$

5.2 Proof of Theorem 2

The proof is once more the same as in [7], with a single change: namely, in (6.19) and (6.20), and in the subsequent definitions such as for the processes $V^{\kappa,n,\theta}$ (notation of that paper), we replace $h_{\kappa,\theta u_n}(a_t, a_t')$ by the following (recall that here we are in the situation where $\kappa = 2$):

$$2\sum_{m=1}^M u_n^{\beta_m-2}y^{\beta_m-2}\chi(\beta_m)a_t^m.$$

All the rest is unchanged, and thus Theorem 2 is proved. Moreover, Theorem 3 is deduced from Theorem 2 as in [7] again.

5.3 Proofs for the rate-efficient estimators

Theorem 3 could be deduced from Theorem 2 as in [7] again. However, we need a sharper result for the iteration procedure of Theorem 4 to work. This method could be applied to other similar situations, but we explain it in our setting. A consequence will be Theorem 4.

We fix the time horizon $T>0$. We are given a number $\rho\in(0,1]$, and M is the biggest integer such that $M\rho\leq 1$. We suppose that for each $u>0$ we have a sequence of estimators $\widetilde{C}_n(u)$, with a specific asymptotic behavior described by the following, where $u_n>0$ is a sequence with $u_n\to 0$ and $u_n/\Delta_n^\varepsilon\to\infty$ for all $\varepsilon>0$, as in (3.6), and where $\mathcal{Y}$ is some finite subset of $(0,\infty)$:

Property (P): *There is a family $\Phi=(\Phi_m:\,1\leq m\leq M)$ of variables on $(\Omega,\mathcal{F},\mathbb{P})$ and two variables $\widetilde{Z}$ and U on an extension of this space, such that if we set*

$$\widetilde{Z}_n(u)=\frac{1}{\sqrt{\Delta_n}}\Big(\widetilde{C}_n(u)-C_T-\sum_{m=1}^M u^{-m\rho}\Delta_n^{m\rho/2}\,\Phi_m\Big) \tag{5.18}$$

for all $u>0$, we have the stable convergence in law

$$\Big(\widetilde{Z}_n(u_n),\Big(\frac{1}{u_n^2}(\widetilde{Z}_n(yu_n)-\widetilde{Z}_n(u_n))\Big)_{y\in\mathcal{Y}}\Big)\stackrel{\mathcal{L}-s}{\longrightarrow}\big(\widetilde{Z},((y^2-1)U)_{y\in\mathcal{Y}}\big). \tag{5.19}$$

To emphasize the ingredients in (P) we sometimes write it as $P(\Phi,\widetilde{Z},U,\mathcal{Y})$. We associate with Φ the sets

$$\Omega(\Phi)_m=\begin{cases}\Omega & \text{if } m=0\\ \{\Phi_1\neq 0\} & \text{if } m=1\\ \{\Phi_1=\cdots=\Phi_{m-1}=0\neq\Phi_m\} & \text{if } 2\leq m\leq M\\ \{\Phi_1=\cdots=\Phi_M=0\} & \text{if } m=M+1\end{cases}$$
$$\Omega'(\Phi)_m=\Omega(\Phi)_m\cap\{\Phi_{m+1}=\cdots=\Phi_M=0\}.$$

For any $\zeta>1$ and $\xi\in\mathbb{R}$ we set

$$\begin{array}{c}G_n(u,\zeta)=\widetilde{C}_n(\zeta u)-\widetilde{C}_n(u),\qquad G'_n(u,\zeta)=\widetilde{C}_n(\zeta^2u)-2\widetilde{C}_n(\zeta u)+\widetilde{C}_n(u)\\ \widetilde{C}'_n(u,\zeta,\xi)=\widetilde{C}_n(u)-\frac{G_n(u,\zeta)^2}{G'_n(u,\zeta)}+u^2\sqrt{\Delta_n}\,\xi.\end{array} \tag{5.20}$$

We then have the following result:

Proposition 9 *Let ζ and ξ and the finite set $\mathcal{Y}$ be given, and set $\mathcal{Y}'=\{y\zeta^j:\,y\in\mathcal{Y},\,j=0,1,2\}$. If the estimators $\widetilde{C}_n(u)$ satisfy $P(\Phi,\widetilde{Z},U,\mathcal{Y}')$, with $U\neq 0$ a.s. on the set $\Omega(\Phi)_{M+1}$, then the estimators $\widetilde{C}'_n(u,\zeta,\xi)$ satisfy $P(\Phi',\widetilde{Z},U',\mathcal{Y})$, where U' is a variable defined on an extension of the space and $\Phi'=(\Phi'_m)_{1\leq m\leq M}$ is a random vector on Ω, such that the following holds for all $m=1,\cdots,M$ and $k=1,\cdots,M+1$:*

$$\begin{array}{l}\Omega(\Phi)_{m-1}\subset\Omega(\Phi')_m,\qquad \Omega'(\Phi)_m\subset\Omega(\Phi')_{M+1}\\ U'=\xi+h_jU\quad\text{on }\Omega(\Phi)_k,\ \text{ where }\ h_k=\begin{cases}\big(\frac{\zeta^{2+k\rho}-1}{\zeta^{k\rho}-1}\big)^2 & \text{if } k\leq M\\ 0 & \text{if } k=M+1.\end{cases}\end{array} \tag{5.21}$$

Here, the variable U' explicitly depends on the numbers ζ and ξ, but the variables Φ'_m will be seen to also depend on ζ.

On the set $\Omega(\Phi)_{M+1}$ we have $\frac{1}{\sqrt{\Delta_n}}\,(\widetilde{C}_n(u_n)-C_T)\stackrel{\mathcal{L}-s}{\longrightarrow}\widetilde{Z}$, and the same holds for $\widetilde{C}'_n(u_n,\zeta,\xi)$, but of course in this case there is no reason to use these new estimators. However, although one might loose some kind of stability for finite samples, asymptotically it does not hurt us to use the new estimators. This is quite an important property, because in practice the variables Φ_m are not observable, so neither is the set $\Omega(\Phi)_{M+1}$.

Proof. 1) We assume $P(\Phi,\widetilde{Z},U)$ with $U\neq 0$ a.s. We simplify our notation by setting

$$\begin{array}{ll}Z_n=\widetilde{Z}_n(u_n), & \widehat{Z}_n^y=\frac{1}{u_n^2}\,(\widetilde{Z}_n(yu_n)-\widetilde{Z}_n(u_n))\\ Y_n^y=\frac{1}{u_n^2}\,(\widetilde{Z}_n(\zeta yu_n)-\widetilde{Z}_n(yu_n))=\widehat{Z}_n^{\zeta y}-\widehat{Z}_n^y & \\ Y_n'^y=\frac{1}{u_n^2}\,(\widetilde{Z}_n(\zeta^2yu_n)-2\widetilde{Z}_n(\zeta yu_n)+\widetilde{Z}_n(yu_n))=\widehat{Z}_n^{\zeta^2y}-2\widehat{Z}_n^{\zeta y}+\widehat{Z}_n^y. & \end{array}$$

Then (P) implies

$$\begin{array}{r}\big(Z_n, (\widehat{Z}_n^y)_{y\in\mathcal{Y}'}, (Y_n^y)_{y\in\mathcal{Y}}, (Y_n'^y)_{y\in\mathcal{Y}}\big) \stackrel{\mathcal{L}-\mathrm{s}}{\longrightarrow} \\ \big(\widetilde{Z}, ((y^2-1)U)_{y\in\mathcal{Y}}, (y^2(\zeta^2-1)U)_{y\in\mathcal{Y}}, (y^2(\zeta^2-1)^2U)_{y\in\mathcal{Y}}\big).\end{array} \tag{5.22}$$

We also write

$$\psi_m = \zeta^{-m\rho} - 1, \qquad w_{n,y} = \frac{\sqrt{\Delta_n}}{yu_n},$$

and we can rephrase the claim as follows: find Φ'_m and U' satisfying (5.21) and such that, if

$$\widetilde{Z}'_n(yu_n) := \frac{1}{\sqrt{\Delta_n}}\Big(\widetilde{C}'_n(yu_n,\zeta) - C_T - \sum_{m=1}^M w_{n,y}^{m\rho}\,\Phi'_m\Big),$$

we have the following stable convergence in law:

$$\big(\widetilde{Z}'_n(u_n), (\overline{Z}_n'^y)_{y\in\mathcal{Y}}\big) \stackrel{\mathcal{L}-\mathrm{s}}{\longrightarrow} \big(\widetilde{Z}, ((y^2-1)\,U')_{y\in\mathcal{Y}}\big). \tag{5.23}$$

2) In view of (5.18) and (5.20), we have

$$\begin{aligned}\widetilde{C}_n(yu_n) &= C_T + \sum_{m=1}^M w_{n,y}^{m\rho}\,\Phi_m + \sqrt{\Delta_n}\,Z_n + u_n^2\sqrt{\Delta_n}\,\widehat{Z}_n^y \\ G_n(yu_n,\zeta) &= \sum_{m=1}^M \psi_m w_{n,y}^{m\rho}\,\Phi_m + u_n^2\sqrt{\Delta_n}\,Y_n^y \\ G'_n(yu_n,\zeta) &= \sum_{m=1}^M \psi_m^2 w_{n,y}^{m\rho}\,\Phi_m + u_n^2\sqrt{\Delta_n}\,Y_n'^y.\end{aligned}$$

We will in fact argue on each set $\Omega(\Phi)_m$ for $m = 1,\cdots,M+1$ separately, since those sets form a partition of Ω. On the set Ω_{M+1} we simply have

$$\widetilde{C}'_n(yu_n,\zeta) = C_T + \sqrt{\Delta_n}\,Z_n + u_n^2\sqrt{\Delta_n}\,\Big(\widehat{Z}_n^y - \frac{(Y_n^y)^2}{Y_n'^y} + y^2\xi\Big)$$

as soon as $Y_n'^y \neq 0$. (5.22) and $U \neq 0$ a.s. imply $\mathbb{P}(Y_n'^y \neq 0) \to 1$, so a simple calculation using (5.22) again shows that (5.23) holds in restriction to Ω_{M+1}, with

$$\Phi'_1 = \cdots = \Phi'_M = 0, \qquad U' = \xi = \xi + h_{N+1}U \quad \text{on } \Omega_{M+1}. \tag{5.24}$$

3) Now, we suppose that we are on the set Ω_m for some $m = 1,\cdots,M$. One easily checks that (with an empty sum set to 0, and since $\Phi_m \neq 0$):

$$\begin{aligned}\widetilde{C}'_n(yu_n,\zeta) &= C_T + \sum_{j=m}^M w_{n,y}^{j\rho}\,\Phi_j + \sqrt{\Delta_n}\,Z_n + u_n^2\sqrt{\Delta_n}\,(\widehat{Z}_n^y + y^2\xi) - \theta_l^{-m\rho} w_{n,y}^{m\rho}\,\Phi_m\,\tfrac{N_n(y)^2}{D_n(y)}, \\ N_n(y) &= 1 + \sum_{j=m+1}^M w_{n,y}^{(j-m)\rho}\,\tfrac{\psi_j\Phi_j}{\psi_m\Phi_m} + u_n^2\sqrt{\Delta_n}\,w_{n,y}^{-m\rho}\,\tfrac{Y_n^l}{\psi_m\Phi_m} \\ D_n(y) &= 1 + \sum_{j=m+1}^M w_{n,y}^{(j-m)\rho}\,\tfrac{\psi_j^2\Phi_j}{\psi_m^2\Phi_m} + u_n^2\sqrt{\Delta_n}\,w_{n,y}^{-m\rho}\,\tfrac{Y_n'^y}{\psi_m^2\Phi_m}.\end{aligned}$$

We make an expansion of the ratio $N_n(y)^2/D_n(y)$, in such a way that we keep all "significant" terms which are of order as big as $u_n^2\sqrt{\Delta_n}$, once multiplied by $w_{n,y}^{m\rho}$. For this, we observe that $w_{n,y} \to 0$ and $w_{n,y}^{j\rho} = \mathrm{o}(u_n^2\sqrt{\Delta_n}\,)$ if and only if $j > M$. Then, after some (tedious) computation, we end up with

$$\widetilde{C}'_n(yu_n,\zeta) = C_T + \sum_{k=1}^5 H_n^y(k) + \sqrt{\Delta_n}\,Z_n + u_n^2\sqrt{\Delta_n}\,\Big(\widehat{Z}_n^y - \frac{2Y_n^y}{\psi_m} + \frac{Y_n'^y}{\psi_m^2} + y^2\xi\Big),$$

where, with $J^m_r(k)$ denoting the set of all r-uples $\{j_i\}$ of integers with $j_i \geq m+1$ and $\sum_{i=1}^r j_i = k$,

$$
\begin{aligned}
H^y_n(1) &= \sum_{j=m+1}^{M} w^{j\rho}_{n,y} \Big(1 - \tfrac{2\psi_j}{\psi_m}\Big)\, \Phi_j \\
H^y_n(2) &= -\sum_{j=m+2}^{M} w^{j\rho}_{n,y} \sum_{(s,k)\in J^m_2(j)} \tfrac{\psi_s\psi_k}{\psi^2_m}\, \tfrac{\Phi_s\Phi_k}{\Phi_m} \\
H^y_n(3) &= \sum_{j=m+1}^{M} w^{j\rho}_{n,y} \sum_{r=1}^{j-m} (-1)^{r+1} \sum_{\{j_i\}\in J^m_r(j+rm-m)} \tfrac{\prod_{i=1}^r(\psi^2_{j_i}\Phi_{j_i})}{\psi^r_m \Phi^{r-1}_m} \\
H^y_n(4) &= 2\sum_{j=m+2}^{M} w^{j\rho}_{n,y} \sum_{r=1}^{j-m-1} (-1)^{r+1} \sum_{k=m+1}^{j-r} \sum_{\{j_i\}\in J^m_r(j+rm-k)} \tfrac{\psi_k\Phi_k \prod_i(\psi^2_{j_i}\Phi_{j_i})}{\psi^{r+1}_m \Phi^{r}_m} \\
H^y_n(5) &= \sum_{j=m+3}^{M} w^{j\rho}_{n,y} \sum_{r=1}^{j-m-2} (-1)^{r+1} \sum_{l=2m+2}^{j+m-r} \sum_{(s,k)\in J^m_2(l)} \sum_{\{j_i\}\in J^m_r(j+rm+m-z)} \tfrac{\psi_s\Phi_s\psi_k\Phi_k \prod(\psi^2_{j_i}\Phi_{j_i})}{\psi^{r+2}_m \Phi^{r+1}_m}.
\end{aligned}
$$

It is then easy, although somehow tedious, to deduce (5.23) in restriction to Ω_m from (5.22), upon taking

$$
\begin{aligned}
& j = 0,\cdots m \quad\Rightarrow\quad \Phi'_j = 0 \\
& j = m+1,\cdots,M \;\Rightarrow\; \Phi'_j = \Big(1 - \tfrac{2\psi_j}{\psi_m}\Big)\,\Phi_j + \sum_{(s,k)\in J^m_2(j)} \tfrac{\psi_s\psi_k}{\psi^2_m}\,\tfrac{\Phi_s\Phi_k}{\Phi_m} \\
&\qquad + \sum_{(s,k)\in J^m_2(j)} \tfrac{\psi_s\psi_k}{\psi^2_m}\,\tfrac{\Phi_s\Phi_k}{\Phi_m} + 2\sum_{r=1}^{j-m-1}(-1)^{r+1} \sum_{k=m+1}^{j-r} \sum_{\{j_i\}\in J^m_r(j+rm-k)} \tfrac{\psi_k\Phi_k\prod_i(\psi^2_{j_i}\Phi_{j_i})}{\psi^{r+1}_m\Phi^r_m} \\
&\qquad + \sum_{r=1}^{j-m-2}(-1)^{r+1} \sum_{l=2m+2}^{j+m-r} \sum_{(s,k)\in J^m_2(l)} \sum_{\{j_i\}\in J^m_r(j+rm+m-z)} \tfrac{\psi_s\Phi_s\psi_k\Phi_k\prod(\psi^2_{j_i}\Phi_{j_i})}{\psi^{r+2}_m\Phi^{r+1}_m} \\
& U' = \xi + \Big(1 - \tfrac{2(\zeta^2-1)}{\psi_m} + \tfrac{(\zeta^2-1)^2}{\psi^2_m}\Big) U = \xi + h_m U.
\end{aligned}
\tag{5.25}
$$

At this stage, if we define Ψ'_m on Ω and U' on the extended space by (5.24) in restriction to Ω_{M+1} and by (5.25) in restriction to Ω_m for any $m = 1,\cdots,M$, and upon using standard properties of the stable convergence in law, we deduce (5.23), whereas (5.21) is obvious. This ends the proof.$\square$

Proof of Theorem 3. We define $\widetilde{C}_n(u)$ to be $\widehat{C}(u)^n_T$, so $\widehat{C}(u,\zeta)^n_T$ is exactly $\widetilde{C}'_n(u,\zeta,0)$, as given by (5.21) with $\xi = 0$. We also set $\rho = 2-\beta_1$, so Theorem 2 implies that $P(\Phi,\widetilde{Z},U)$ holds with $\Phi_1 = A^1_T$ and $\Phi_2 = \cdots = \Phi_M = 0$ and with $\widehat{Z} = Z_T$ and $U = \overline{Z}_T$. Note that $U \neq 0$ a.s. because $C_T > 0$ a.s. by assumption. Theorem 2 and the previous proposition yield that the estimators $\widetilde{C}'_n(u,\zeta,0)$ satisfy $P(0,\widetilde{Z},U',\{1\})$, and (5.19) gives us the result. $\square$

Proof of Theorem 4. 1) We assume (B) here, with $\rho > \rho_0$ and $\rho_0 \in [0,1)$, so we have $\beta_m = 2-\nu_m\rho$ for integers $1 \leq \nu_1 < \cdots < \nu_M \leq 1/\rho$.

We start with some simple changes in the setting. First, in the de-biasing sum in (3.5) one can drop out all terms such that $\beta_m < 1$, because those terms multiplied by $1/\sqrt{\Delta_n}$ and even by $1/u^2_n\sqrt{\Delta_n}$, go to 0, hence can be removed without altering Theorem 2. In other words, we may assume $\beta_m \geq 1$ for all m. Observe that, now, all ν_m defined above are at most equal to biggest integer M' such that $M'\rho_0 \leq 1$, because $\rho \geq \rho_0$.

Second, we add "fictitious" indices, so that the indices fill in the whole set $\{2-m\rho : m = 1,\ldots,M'\}$, and we set the associated process a^m to be identically 0 for all those fictitious indices. This does not affect the de-biasing term in (3.5), so Theorem 2 stays valid with these new indices. Therefore, without loss of generality we may and will assume that we have $\beta_m = 2-m\rho$ for all $m \leq M$ (observe that now we may again have $\beta_m < 1$ for some values of m, but we will no longer suppress those).

2) We will prove by induction on j the following, where $\mathcal{Y}_m$ is the set $\{\zeta^j : j = 0,1,\ldots,2m+2\}$:

Claim: The estimators $\widehat{C}(u,\zeta,j)^n_T$ satisfy $P(\Phi^j,\widetilde{Z},U^j,\mathcal{Y}_{M-j}j)$; we have $\Phi^j_1=\cdots=\Phi^j_{j\wedge M}=0$; furthermore U^j is $\mathcal{F}$-conditionally Gaussian non-degenerate if $j=0$ or in restriction to the complement of the set $\Omega(\Phi^{j-1})_{M+1}$ if $j\geq 1$, and $U^j=1$ otherwise.

When $j=0$, this claim is simply Theorem 2 with $\mathcal{Y}=\mathcal{Y}_{2M}$, upon taking

$$\Phi^0_m=A^m_T,\qquad \widetilde{Z}=Z_T,\qquad U^0=\overline{Z}_T, \tag{5.26}$$

and U^0 is $\mathcal{F}$-conditionally Gaussian non-degenerate because $C_T>0$ a.s.

Suppose now the claim holds for some $j\geq 0$. We will apply Proposition 9 to $\widetilde{C}_n(u,\zeta,j)=\widehat{C}(u,\zeta,j)^n_T$. Indeed, the associated estimators in (5.20) are exactly $\widehat{C}(u,\zeta,j+1)^n_T$, provided we take $\xi\equiv 1$. Note also that $U^j\neq 0$ a.s. on Ω because our claim holds for j. Then this proposition tells us that the claim for $j+1$ holds (use (5.21) to obtain that $\Phi^{j+1}_1=\cdots=\Phi^{j+1}_{(j+1)\wedge M}=0$, and also that $U^{j+1}=\xi=1$ on the set $\Omega(\Phi^{j-1})_{M+1}$ and that U^{j+1} is $\mathcal{F}$-conditionally Gaussian non-degenerate on the complement $\Omega(\Phi^{j-1})^c_{M+1}$, because the numbers h_k are positive for all $k\leq M$).

3) What precedes shows that the claim holds for all j, up to $M+1$. Now, M is the number N of iterations made in the theorem, that is, $\widetilde{C}_n(u,\zeta,M+1)=\widehat{C}(u,\zeta,N)^n_T$. On the other hand, if $j=M+1$ all components of Φ^j vanish identically. Hence the estimators $\widehat{C}(u,\zeta,N)^n_T$ satisfy $P(0,\widetilde{Z},U^N,\{1\})$, and we recall that $\widetilde{Z}$ is given in (5.26). This gives us the result. $\square$

References

[1] Ait-Sahalia, Y. and J. Jacod (2008). Fisher's information for discretely sampled Lévy processes. *Econometrica*, **76**, 727-761.

[2] Ait-Sahalia, Y. and J. Jacod (2014). *High-Frequency Financial Econometrics.* Princeton University Press.

[3] Barndorff-Nielsen, O.E. and N. Shephard (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *J. of Financial Econometrics* **2**, 1–48.

[4] Barndorff-Nielsen, O.E., N. Shephard and M. Winkel (2006). Limit Theorems for Multipower Variation in the Presence of Jumps. *Stochastic Processes and Their Applications* **116**, 796-806.

[5] Clément, E, S. Delattre and A. Gloter (2013). An Infinite Dimensional Convolution Theorem with Applications to the Efficient Estimation of the Integrated Volatility. *Stochastic Processes and Their Applications* **123**, 2500-2521.

[6] Jacod, J. and P. Protter (2012). *Discretization of Processes*, Springer-Verlag, Berlin.

[7] Jacod, J. and V. Todorov (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Annals of Statistics* **42**, 1029-1069.

[8] Jacod, J. and M. Reiss (2014). A Remark on the Rates of Convergence for the Integrated Volatility Estimation in the Presence of Jumps. *Annals of Statistics* **42**, 1131–1144.

[9] Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari* **LXIV**, 19–47.

[10] Mancini, C. (2011). The Speed of Convergence of the Threshold Estimator of Integrated Variance. *Stochastic Processes and their Applications* **121**, 845-855.

[11] Todorov, V. and G. Tauchen (2012). The Realized Laplace Transform of Volatility. *Econometrica* **80**, 1105–1127.

[12] Vetter, M. (2010). Limit theorems for bipower variation of semimartingales.*Stochastic Processes and Their Applications* **120**, 22-38.