Scandinavian Journal of Statistics, Vol. 36: 270–296, 2009
doi: 10.1111/j.1467-9469.2008.00622.x

Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps

CECILIA MANCINI

Università di Firenze, Dipartimento di matematica per le decisioni

ABSTRACT. We consider a stochastic process driven by diffusions and jumps. Given a discrete record of observations, we devise a technique for identifying the times when jumps larger than a suitably defined threshold occurred. This allows us to determine a consistent non-parametric estimator of the integrated volatility when the infinite activity jump component is Lévy. Jump size estimation and central limit results are proved in the case of finite activity jumps. Some simulations illustrate the applicability of the methodology in finite samples and its superiority on the multipower variations especially when it is not possible to use high frequency data.

Key words: asymptotic properties, discrete observations, integrated volatility, models with stochastic volatility and jumps, non-parametric estimation, threshold

1. Introduction

In financial mathematics, the most commonly used models to represent asset prices or interest rates are in the class of the stochastic processes X starting from an initial value $x_0 \in \mathbb{R}$ at time $t=0$ and being such that

$$dX_t = a_t\,dt + \sigma_t\,dW_t + dJ_t, \quad t \in]0,T], \tag{1}$$

where a and σ are progressively measurable processes, W is a standard Brownian motion and J is a pure jump process. Given discrete observations, both for pricing and hedging aims and for financial econometrics applications, it is important to separate the contributions of the diffusion part and jump part of X (see Andersen *et al.*, 2005; Barndorff-Nielsen & Shephard, 2007; Cont *et al.*, 2007).

A jump process is said to have *finite activity* (FA) when it makes a.s. a finite number of jumps in each finite time interval, otherwise it is said to have *infinite activity* (IA). In general, a jump process is the sum of an FA and an IA component. We provide an estimate of the *integrated volatility* $\int_0^T \sigma_t^2\,dt$, denoted by IV, given discrete observations $\{x_0, X_{t_1}, \ldots, X_{t_n}\}$, which is consistent when either J has FA or the IA component of J is Lévy. We make use of a suitably defined threshold. When J has FA we also give an estimate of jump times and sizes, while when J has IA we can identify the instants when jumps larger than the threshold occurred.

The method we propose here extends previous work in Mancini (2001) and Mancini (2004), allowing for infinite jump activity and very mild assumptions on the processes a and σ. An alternative extension of the threshold estimation method has been made in Jacod (2008), where, in order to obtain a central limit theorem, the volatility dynamics has to be specified.

It has been shown that diffusion models are not any longer adequate for describing the behaviour of many financial assets (see for example Andersen *et al.*, 2002; Das, 2002; Eraker *et al.*, 2003; Barndorff-Nielsen & Shephard, 2006; Aït-Sahalia & Jacod, 2008). In the literature on non-parametric inference for stochastic processes driven by diffusions plus jumps, several approaches have been proposed to separate the diffusion part and the jump part given discrete observations.

Berman (1965) defined *power variation* estimators of the sum of given powers of the jumps. Recently these have been recovered and developed in Barndorff-Nielsen & Shephard (2004), Woerner (2006) and Jacod (2008).

Barndorff-Nielsen & Shephard (2004, 2006) define and use the *bipower* and the *multipower variation* processes (MPV) to estimate $\int_0^T \sigma_t^p \, dt$ for given values of p, and in particular they focus on $p=2$. In these works they assume that σ is independent of the leading Brownian motion (*no leverage* assumption) and that the jump process has finite activity. In particular, they build a test for the presence of jumps in the data-generating process. Barndorff-Nielsen *et al.* (2006b) and Woerner (2006) show that, in specified cases, the consistency and the central limit theorem for the multipower variation estimators can be extended in the presence of infinite activity jump processes. Combining the results in Barndorff-Nielsen *et al.* (2006a) and (2006b), we find that the extension is possible even under leverage, where a dynamics for σ has to be specified.

Bandi & Nguyen (2003) and Johannes (2004) assume that $a_t \equiv a(X_t), \sigma_t \equiv \sigma(X_t)$ and that J has FA bounded jumps. They use Nadaraya–Watson kernels to obtain pointwise estimators of the functions $a(x)$ and $\sigma(x)$ and aggregate information about J. Mancini & Renò (2006) combine the kernel and the threshold methods to improve the estimation of the jump part and extend the results to the infinite jump activity framework.

The advantages of the threshold technique can be summarized as follows. First, in the FA case, threshold estimation is a more effective way to identify intervals $]t_{j-1}, t_j]$ where J jumped. Second, the threshold estimator $\widehat{\mathrm{IV}}_h$ of IV is efficient (in the Cramér–Rao inequality lower bound sense) while any multipower variation estimator is not. This turns out to be important, since a comparison on simulations with the multipower variation estimators shows that $\widehat{\mathrm{IV}}_h$ gives more reliable results, especially when the frequency of the observations is not very high. Third, the consistency of the threshold estimator holds whatever be the dynamics for σ (even under leverage) and when the observations are not equally spaced. The good performance of our estimator on finite samples with both large and small steps h between the observations is shown within three different simulated models, which are common in finance.

An outline of the paper is as follows: In section 2 we introduce the framework and the notation; in section 3 we deal with the case where J has FA: we show that by the threshold method we can asymptotically identify each instant of jump. As a consequence we obtain threshold estimators of $\int_0^T \sigma_s^2 \, ds$ and of each stochastic size of the occurred jumps. Using results in Barndorff-Nielsen *et al.* (2006a) and in Barndorff-Nielsen & Shephard (2007) we show the asymptotic normality of $\widehat{\mathrm{IV}}_h$, whatever be the dynamics for σ. Moreover we find the asymptotic distribution of the estimation error of the sizes of jumps under the no leverage assumption and when the jump component is a compound Poisson process. Section 4 is devoted to the case when the underlying process contains an infinite activity Lévy jump part: in a quite simple way we show that the threshold estimator of IV is still consistent, even under leverage and when the observations are not equally spaced. The proofs are deferred to the appendix. Section 5 shows the performance of the estimator of IV in finite samples within three different simulated models and a comparison with the multipower variations performances. Section 6 concludes.

2. The framework

On the filtered probability space $(\Omega, (\mathcal{F}_t)_{t\in[0,\,T]}, \mathcal{F}, \mathrm{P})$, let W be a standard Brownian motion and J be a pure jump process given by $J_1 + \tilde{J}_2$, where J_1 has FA and $\tilde{J}_2$ is an IA Lévy pure jump process. Let $(X_t)_{t\in[0,\,T]}$ be a real process starting from $x_0 \in \mathbb{R}$ and evolving in time as in (1), where a and σ are progressively measurable processes which guarantee that (1) has a unique strong solution on $[0, T]$ which is adapted and right continuous with left limits (càdlàg). For results on existence and uniqueness of solution of (1), see for example Ikeda & Watanabe (1981) and Protter (1990). Suppose that on the finite and fixed time horizon $[0, T]$ we dispose of a discrete record $\{x_0, X_{t_1}, \dots, X_{t_{n-1}}, X_{t_n}\}$ of $n+1$ observations of a realization of X. For simplicity we consider the case of equally spaced observations with $t_i = ih$, for a given lag h, so that $T = nh$. This simplification is not essential, as it will be remarked later.

When J is a Lévy process, we can always decompose it as the sum of the jumps larger than one and of the compensated jumps smaller than one, as follows:

$$J = J_1 + \tilde{J}_2, \quad J_{1s} := \int_0^s \int_{|x|>1} x\mu(\mathrm{d}t, \mathrm{d}x), \quad \tilde{J}_{2s} := \int_0^s \int_{|x|\le 1} x(\mu(\mathrm{d}t, \mathrm{d}x) - \nu(\mathrm{d}x)\,\mathrm{d}t), \tag{2}$$

where μ is the Poisson random measure of the jumps of J, $\tilde{\mu}(\mathrm{d}t, \mathrm{d}x) = \mu(\mathrm{d}t, \mathrm{d}x) - \nu(\mathrm{d}x)\,\mathrm{d}t$ is the compensated measure and ν is the Lévy measure of J (see Ikeda & Watanabe, 1981 or Sato, 1999). $\tilde{J}_2$ is a square integrable martingale with infinite activity of jump. For each s, $\mathrm{var}(\tilde{J}_{2s}) = s\int_{|x|\le 1} x^2 \nu(\mathrm{d}x) := s\sigma^2(1) < \infty$. J_1 is a compound Poisson process with finite activity of jump, and we can also write $J_{1s} = \sum_{j=1}^{N_s} \gamma_j$, where N is a Poisson process with constant intensity λ, jumping at times denoted by $(\tau_j)_{j=1..N_T}$, and each γ_j, also denoted γ_{τ_j}, is the size of the jump occurred at τ_j. The random variables γ_j are i.i.d. and independent of N.

We consider slightly more general jump processes $J = J_1 + \tilde{J}_2$, where $\tilde{J}_2$ is an IA Lévy pure jump process and J_1 is a general FA jump process, that is it has the form $J_{1s} = \sum_{j=1}^{N_s} \gamma_j$, where N is a non-explosive counting process with not necessarily constant intensity, and the real random variables γ_j are not necessarily i.i.d., nor independent of N.

Denote by $\tau^{(i)}$ the first instant a jump occurs within $]t_{i-1}, t_i]$, if $N_{t_i} - N_{t_{i-1}} \ge 1$; by $\gamma^{(i)}$ the size of this first jump within $]t_{i-1}, t_i]$, if $N_{t_i} - N_{t_{i-1}} \ge 1$; by $\underline{\gamma} := \min_{j=1..N_T} |\gamma_j|$.

The next section deals with the case where J has FA, i.e. $\tilde{J}_2 \equiv 0$, while in section 4 we allow J to have infinite activity, where $\tilde{J}_2$ is Lévy.

We use the following further notation throughout the paper:

- For any semimartingale Z, let us denote by $\Delta_i Z$ the increment $Z_{t_i} - Z_{t_{i-1}}$ and by ΔZ_t the size $Z_t - Z_{t-}$ of the jump which (possibly) occurred at time t.
- $[Z]$ is the quadratic variation process associated to Z, $[Z^{(h)}]_T$ is the estimator $\sum_{i=1}^n (\Delta_i Z)^2$ of the quadratic variation $[Z]_T$ at time T.
- $H \cdot W$ is the process given by the stochastic integral $\int_0^{\cdot} H_s\,\mathrm{d}W_s$.
- $\mathrm{IV}_t := \int_0^t \sigma_u^2\,\mathrm{d}u$ is the *integrated volatility*. $\mathrm{IQ}_t := \int_0^t \sigma_u^4\,\mathrm{d}u$ is called *integrated quarticity* of X in the econometric literature (see for example Barndorff-Nielsen & Shephard, 2006).
- By c (lower case) we denote generically a constant.
- Plim means 'limit in probability'; dlim means 'limit in distribution'; $\xrightarrow{st}$ denotes stable convergence in law.
- If η is an r.v., $M\mathcal{N}(0, \eta^2)$ indicates the mixed Gaussian law having characteristic function $\phi(\theta) = E[\mathrm{e}^{-\frac{1}{2}\eta^2\theta^2}]$.

3. Finite activity jumps

3.1. Consistency

An important variable related to X and containing the quantity IV we want to estimate is the quadratic variation at T,

$$[X]_T = \int_0^T \sigma_t^2\,\mathrm{d}t + \int_0^T \int_{\mathbb{R}} x^2 \mu(\mathrm{d}x, \mathrm{d}t). \tag{3}$$

As is known, an estimate of $[X]_T$ is given by $\sum_{i=1}^{n}(\Delta_i X)^2$. We consider in this section the case in which J has FA, so that (3) becomes

$$[X]_T = \int_0^T \sigma_t^2\,\mathrm{d}t + \sum_{j=1}^{N_T} \gamma_{\tau_j}^2,$$

and the quadratic variation gives us only an aggregate information containing both IV and the squared jump sizes. In order to isolate the contribution of $\int_0^T \sigma_t^2\,\mathrm{d}t$ to $[X]_T$, the key point here is to exclude the time intervals $]t_{i-1}, t_i]$ where J jumped. The following theorem provides an instrument to asymptotically identifying such intervals.

Theorem 1

Identification of the intervals where no jumps occurred.

Suppose that $J = \sum_{j=1}^{N_t} \gamma_j$ is a finite activity jump process where N is a non-explosive counting process and the random variables γ_j satisfy, $\forall t \in [0, T]$, $P\{\Delta N_t \neq 0, \gamma_{N_t} = 0\} = 0$. Suppose also that

(1) a.s. $\limsup_{h\to 0} \dfrac{\sup_{i\in\{1,\ldots,n\}} |\int_{t_{i-1}}^{t_i} a_s\,\mathrm{d}s|}{\sqrt{2h\log\frac{1}{h}}} \le C(\omega) < \infty,$

(2) a.s. $\limsup_{h\to 0} \dfrac{\sup_i |\int_{t_{i-1}}^{t_i} \sigma_s^2\,\mathrm{d}s|}{h} \le M(\omega) < \infty,$

(3) $r(h)$ is a deterministic function of the lag h between the observations, such that $\lim_{h\to 0} r(h) = 0$, and $\lim_{h\to 0}(h\log\frac{1}{h})/r(h) = 0$.

Then, for P-almost all ω, $\exists \bar{h}(\omega) > 0$ such that $\forall h \le \bar{h}(\omega)$ we have

$$\forall i = 1, \ldots, n, \quad I_{\{(\Delta_i X)^2 \le r(h)\}}(\omega) = I_{\{\Delta_i N = 0\}}(\omega). \tag{4}$$

Assumption 3 indicates how to choose the *threshold* $r(h)$. The absolute value of the increments of any path of the Brownian motion (and thus of a stochastic integral with respect to the Brownian motion) tends a.s. to zero at the same speed as the deterministic function $\sqrt{2h\ln\frac{1}{h}}$. Therefore, for small h, when we find that the squared increment $(\Delta_i X)^2$ is larger than $r(h) > 2h\ln\frac{1}{h}$, then it is likely that some jumps occurred.

Remarks.

(i) $r(h) = h^{\beta}$ is a possible choice for the threshold for any $\beta \in]0, 1[$, since it satisfies assumption 3 of the theorem.

(ii) Mancini & Renò (2006) show that it is possible to consider also a time varying threshold.

(iii) Assumptions 1 and 2 are satisfied if the paths $(a_s(\omega))_s$, $(\sigma_s(\omega))_s$ are a.s. bounded on $[0, T]$. In particular they are satisfied as soon as a and σ have càdlàg paths.

(iv) The condition that, $\forall t \in [0, T]$, $P\{\Delta N_t \neq 0, \gamma_{N_t} = 0\} = 0$ means that a.s. when a jump occurs the size has to be non-zero, otherwise we cannot recognize it. Note that any FA Lévy process satisfies the condition, since $\nu\{0\} = 0$. For example, this is the case for a compound Poisson process with Gaussian sizes of jump.

(v) Frequently, in practice, the lag $\Delta t_i := t_i - t_{i-1}$ between the observations of an available record $\{x_0, X_{t_1}, \ldots, X_{t_{n-1}}, X_{t_n}\}$ is not constant (not equally spaced observations). If we take $h := \max_i \Delta t_i$, then theorem 1 and corollary 1 below are still valid as they are stated (see the remark in Appendix after the proof of theorem 1).

Alternatively, it is asymptotically equivalent to directly compare each $(\Delta_i X)^2$ with the relative $r(\Delta t_i)$. In this case, theorem 1 is modified as follows (see the remark after the proof in the Appendix).

Corollary 1

Identification of the intervals where no jumps occurred, when observations are not equally spaced.

Under the same assumptions on the jump process J as in theorem 1, set $h = \max_i \Delta t_i$. *If*

(1) a.s. $\limsup_{h\to 0}\sup_i \frac{|\int_{t_{i-1}}^{t_i} a_s\,\mathrm{d}s|}{\sqrt{2\Delta t_i \log \frac{1}{\Delta t_i}}} \le C(\omega) < \infty$,

(2) a.s. $\limsup_{h\to 0}\sup_i \frac{|\int_{t_{i-1}}^{t_i} \sigma_s^2\,\mathrm{d}s|}{\Delta t_i} \le M(\omega) < \infty$,

(3) $r:\mathbb{R}\to\mathbb{R}$ *is a function, such that* $\lim_{u\to 0} r(u) = 0$ *and* $\lim_{u\to 0}(u \log \frac{1}{u})/r(u) = 0$,

then, for P-almost all ω, $\exists \bar{h}(\omega) > 0$ *such that* $\forall h \le \bar{h}(\omega)$ *we have*

$$\forall i = 1, \dots, n, \quad I_{\{(\Delta_i X)^2 \le r(\Delta t_i)\}} = I_{\{\Delta_i N = 0\}}.$$

Let us return to equally spaced observations. Define

$$\widehat{\mathrm{IV}}_h = \sum_{i=1}^n (\Delta_i X)^2 I_{\{(\Delta_i X)^2 \le r(h)\}}.$$

By virtue of theorem 1, a.s., for small h, $\widehat{\mathrm{IV}}_h$ includes only the squared increments $(\Delta_i X)^2$ relative to those intervals $]t_{i-1}, t_i]$ where no jumps occurred. Since a.s. only $N_T < \infty$ terms are excluded, we can show that $\widehat{\mathrm{IV}}_h$ has the same asymptotic behaviour as $\sum_{i=1}^n (\int_{t_{i-1}}^{t_i} a_s\,\mathrm{d}s + \int_{t_{i-1}}^{t_i} \sigma_s^2\,\mathrm{d}W_s)^2$, which tends in probability to IV, and thus we obtain that $\{\widehat{\mathrm{IV}}_h\}_h$ converges to IV. This result is stated in the next corollary.

Corollary 2

Under the assumptions of theorem 1 we have $\mathrm{P}\lim_{h\to 0} \widehat{\mathrm{IV}}_h = \mathrm{IV}$.

3.2. Central limit theorems

As a corollary of theorem 2.2 in Barndorff-Nielsen *et al.* (2006a, see the Appendix), from our theorem 1 we obtain a threshold estimator of $\int_0^T \sigma_t^4\,\mathrm{d}t$, which is an alternative to the power variation. An estimate of $\int_0^T \sigma_t^4\,\mathrm{d}t$ is needed in order to give the asymptotic law of the approximation error $\widehat{\mathrm{IV}}_h - \mathrm{IV}$. We obtain a central limit result (CLT) for $\widehat{\mathrm{IV}}_h$ whatever be the dynamics for σ.

Proposition 1

Under the same assumptions on the jump process J and the same condition (3) as in theorem 1, if the processes a and σ *are càdlàg, we have that as* $h \to 0$,

$$\mathrm{I\hat{Q}}_h := \frac{1}{3h}\sum_{i=1}^n (\Delta_i X)^4 I_{\{(\Delta_i X)^2 \le r(h)\}} \xrightarrow{P} \int_0^T \sigma_t^4\,\mathrm{d}t.$$

Finally, as a corollary of theorem 1 in Barndorff-Nielsen & Shephard (2007, see the Appendix) we have the following result of asymptotic normality for our estimator $\widehat{\mathrm{IV}}_h$.

Theorem 2

Under the same assumptions on the jump process J and the same condition (3) as in theorem 1, if a and $\sigma \not\equiv 0$ *are càdlàg processes, then we have* $(\widehat{\mathrm{IV}}_h - \mathrm{IV})/\sqrt{2h\widehat{IQ}_h} \xrightarrow{d} \mathcal{N}(0,1)$.

By theorem 1 and by the fact that, for small h, the probability of more than one jump over an interval $]t_{i-1}, t_i]$ is low, it is clear that estimators of the jump instants are given by $\hat{\tau}^{(1)} = \inf\{t_i \geq 0 : (\Delta_i X)^2 > r(h)\}$ and, for $j > 1$, $\hat{\tau}^{(j)} = \inf\{t_i > \hat{\tau}^{(j-1)} : (\Delta_i X)^2 > r(h)\}$. Moreover, a natural estimate of each realized jump size is given by

$$\hat{\gamma}^{(i)} := \Delta_i X I_{\{(\Delta_i X)^2 > r(h)\}},$$

since when a jump occurs, then the contribution of $\int_{t_{i-1}}^{t_i} a_u \,\mathrm{d}u + \int_{t_{i-1}}^{t_i} \sigma_u \,\mathrm{d}W_u$ to $\Delta_i X$ is asymptotically negligible. In Mancini (2004) we have shown the consistency of each $\hat{\gamma}^{(i)}$ when $T \to \infty$. Here, for fixed $T < \infty$, we make it more precise, showing that under the (restrictive) no leverage assumption and when $J \equiv J_1$ is Lévy, the speed of convergence is $\sqrt{h}$.

Theorem 3

If the following conditions are met:

(1) J is a compound Poisson process,

(2) a.s. $\limsup_{h \to 0} \frac{\sup_i |\int_{t_{i-1}}^{t_i} a_s \,\mathrm{d}s|}{h^{\mu}} \leq C(\omega) < \infty$ *for some* $\mu > 0.5$ *(which is the case if a is càdlàg),*

(3) σ *is an adapted stochastic process with continuous paths and is independent of W and N, with* $E[\int_0^T \sigma_s^2 \,\mathrm{d}s] < \infty$,

(4) the threshold $r(h)$ *is chosen as in theorem 1,*

then

$$\sqrt{n} \sum_{i=1}^{n} \left(\hat{\gamma}^{(i)} - \gamma^{(i)} I_{\{\Delta_i N \geq 1\}} \right) \xrightarrow{d} M\mathcal{N}\left(0, T \int_0^T \sigma_s^2 \,\mathrm{d}N_s \right).$$

3.3. Comparison with the multipower variations estimators

The advantages of the non-parametric threshold method are at least three.

The threshold estimator of IV is efficient (in the Cramér–Rao inequality lower bound sense, see Aït-Sahalia, 2004, for constant σ), since $(\widehat{\mathrm{IV}}_h - \mathrm{IV})/\sqrt{h\widehat{\mathrm{IQ}}_h}$ tends in distribution to $\mathcal{N}(0, 2)$. In contrast, Barndorff-Nielsen & Shephard (2006, p. 29) show that, when X is a diffusion, for the bipower variation $V_{1,1}(X)$ (see (7) for the definition) we have that $(\mu_1^{-2} V_{1,1}(X) - \mathrm{IV})/\sqrt{h\mathrm{IQ}}$ tends in distribution to $\mathcal{N}(0, \frac{\pi^2}{4} + \pi - 3)$.

Note that while the bipower variation can consistently estimate IV even in presence of jumps, a CLT for $V_{1,1}(X)$ holds only when X is a diffusion (Barndorff-Nielsen & Shephard, 2006; Barndorff-Nielsen *et al.*, 2006a; Woerner, 2006). When there are jumps, multipower variations $V_{r_1,\ldots r_k}(X)$ can consistently estimate $\int_0^T |\sigma_s|^{\sum_{j=1}^k r_j} \,\mathrm{d}s$; however, no central limit theorem (with speed of convergence $\sqrt{h}$) is guaranteed unless $\max_i r_i < 1$ (see section 5), so that, to estimate IV, multipower variations $V_{r_1,\ldots r_k}(X)$ with $k \geq 3$ are needed. For example, in the framework of this paper, $V_{\frac{1}{3},\frac{1}{3}}(X)$ gives an asymptotically Gaussian estimator of $\int_0^T \sigma_s^{2/3} \,\mathrm{d}s$. In fact, just to make an example, we present a simulation study even on $V_{\frac{1}{3},\frac{1}{3}}(X)$, since using it we obtain an asymptotic variance of $C_{\frac{1}{3},\frac{1}{3}} \mu_{\frac{1}{3}}^{-4} = 0.37$ ($C_{\frac{1}{3},\frac{1}{3}}$ is given below (8)), which is much less than the asymptotic variances of the tripower variation estimators we use below. $V_{\frac{1}{3},\frac{1}{3}}(X)$ is not really pertinent here, since an estimate of $\int_0^T \sigma_s^{2/3} \,\mathrm{d}s$ is not directly comparable with that of IV. Unless σ is constant, to pass to an estimate of $\int_0^T \sigma_s^2 \,\mathrm{d}s$, a differentiation and then another integration would be necessary, which can introduce new bias. However, the simulations presented show that even before making the transformation the finite sample performance of $V_{\frac{1}{3},\frac{1}{3}}(X)$ is worse than $\widehat{\mathrm{IV}}_h$ when the sampling frequency is not very high.

Returning to the estimation of IV, if we want to use MPVs, the theory leads us to use at least three powers r_i.

Note that, at least if we choose $r_1 = \cdots = r_k$, the inefficiency of $V_{r_1,\ldots,r_k}(X)$ increases as k increases. So to estimate IV using MPV it is convenient to use tripower variation. Considering equal powers r_i is now of common use (see for example Huang & Tauchen, 2005). For $V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}}(X)$ the asymptotic variance is 3.06.

Alternatively, in the simulation study we try to use also tripower variations $V_{r_1,r_2,r_3}(X)$ with powers r_i such that r_1 and r_2 are close to 1 and $\sum_{i=1}^{3} r_i = 2$. In fact, in this case, there is an improvement in efficiency. The infimum of the asymptotic variance, as r_1 and r_2 are close to 1, is $\frac{\pi^2}{4} + \pi - 3 = 2.609$. However, this small degree of inefficiency of tripower variation is quite important in the applications, as the presented figures and tables show.

The inefficiency is in fact very important when h is 'large' (i.e. $h \in [1/21{,}000; 1/252]$, corresponding, in a 7 hours open market, to a time lag between observations ranging from 5 minutes up to 1 day) or when the volatility is very small. The figures of this paper show that for $n = 1000$ and $h = 1/n$ the multipower variation estimators perform much worse than the threshold estimator. If we decrease the step h between observations to $1/(252 \times 288)$ and take $n = 288$ (as Barndorff-Nielsen & Shephard, 2006, do), then the multipower variation estimators improve, i.e. these latter estimators need a very small h and a time horizon T much smaller than 1 year. However, it is important to reach reliable results even for 'large' values of h: firstly this avoids all problems connected with microstructure noises which are present in ultra high frequency data (usually for values of h below 1/21,000 in a 7 hours open market); secondly it is possible to apply the threshold technique even to those assets, or commodities prices, for which observations are available not so frequently.

In Cont & Mancini (2008) estimation of IV is applied to a jump-diffusion process with given small constant volatility v^2 to test for the presence of a diffusion component in a data-generating process. We verify on simulations that the use of the threshold estimator gives reliable test results, while the use of multipower variations does not.

The second advantage of the threshold technique is that, in the case of an FA jump component, a.s. for $h < \bar{h}(\omega)$ exact identification of the location of the jumps (and estimation of the sizes) is possible, with relative evaluation of the speed of convergence of the estimators (see even Mancini, 2004). On the contrary, this has not been done up to now using MPVs. This property of the threshold method has important consequences. For example, we can adapt known estimation methods for diffusion processes also to jump-diffusion processes. In Mancini & Renò (2006), to estimate non-parametrically the coefficients of a jump-diffusion model, a kernel method is applied after having removed the jump intervals. To show the consistency of the estimators some work is necessary, but it is crucial to be able to identify the intervals where some jumps occurred.

Third advantage is that a CLT for the threshold estimator holds whatever be the dynamics for σ (even under leverage) and when the observations are not equally spaced.

4. Infinite activity jumps

Let us now consider the case when J has possibly infinite activity. Denote

$$X_{0s} := \int_0^s a_t \,\mathrm{d}t + \int_0^s \sigma_t \,\mathrm{d}W_t, \quad X_1 := X_0 + J_1, \tag{5}$$

and note that since $\int_{|x|\le 1} x^2 \nu(\mathrm{d}x) < +\infty$, then $\sigma^2(\epsilon) := \int_{|x|\le\epsilon} x^2 \nu(\mathrm{d}x) \to 0$ as $\epsilon \to 0$.

In fact our threshold estimator is still able to isolate IV through the observed data. The reason is that now we have

$$[X]_T = \int_0^T \sigma_u^2\,\mathrm{d}u + \int_0^T \int_{|x|>0} x^2 \mu(\mathrm{d}x, \mathrm{d}u) = \int_0^T \sigma_u^2\,\mathrm{d}u + \sum_{s\le T} (\Delta J_{1s})^2 + \sum_{s\le T} (\Delta \tilde{J}_{2s})^2,$$

and for all $\delta > 0$ the threshold $r(h)$ cuts off all the jumps of J_1 and the jumps of $\tilde{J}_2$ larger, in absolute value, than $\sqrt{\delta + 4r(h)}$. However as $r(h) \to 0$, since δ is arbitrary, then every jump of $\tilde{J}_2$ is cut off.

Theorem 4

Let the assumption 1 (pathwise boundedness condition on (a), assumption 2 (pathwise boundedness condition on σ) and assumption 3 (choice of the function $r(h)$) of theorem 1 hold. Let $J = J_1 + \tilde{J}_2$ be such that J_1 has FA with $P\{\Delta_i N \neq 0\} = O(h)$ as $h \to 0$, for all $i = 1..n$; let $\tilde{J}_2$ be Lévy and independent of N. Then $\operatorname*{Plim}_{h\to 0} \widehat{\mathrm{IV}}_h = \mathrm{IV}$.

Remarks.

(i) The result is still valid if we have not equally spaced observations so long as we set $h := \max_i \Delta t_i$ (see the remark in the Appendix after the proof of theorem 4).

(ii) Jacod (2008) proves the consistency of the threshold estimator when J is a more general pure jump semimartingale, with the choice $r(h) = ch^{\beta}$. The proof we present here is simpler and it allows to understand the contribution of the different jump terms to the (asymptotically negligible) estimation bias. Moreover, our approach allows to prove a central limit theorem for $\widehat{\mathrm{IV}}_h$ only assuming that σ is càdlàg, while in Jacod (2008) an assumption on the dynamics of σ is needed in order to get a CLT. This topic is further developed in Cont & Mancini (2008).

(iii) Consistently with the results in Jacod (2008), and in Barndorff-Nielsen *et al.* (2006b) for the multipower variations, the asymptotic normality at speed $\sqrt{h}$ of our estimator of $\int_0^T \sigma_s^2\,\mathrm{d}s$ does not hold in general if X has an infinite activity jump component. Namely, in Cont & Mancini (2008) we find that the speed of convergence of $\widehat{\mathrm{IV}}_h$ is $\sqrt{h}$ when the jump component has a moderate jump activity (when the Blumenthal–Getoor index α of J belongs to $[0, 1[$), while the speed is less than $\sqrt{h}$ if the activity of jump of J is too wild ($\alpha \in [1, 2[$).

(iv) Due to its good properties, the threshold technique (TC) has a variety of applications. As previously remarked, Mancini & Renò (2006) combine it with a kernel method to obtain non-parametric estimation of $a(x)$ and $\sigma(x)$ when $a_t \equiv a(X_t)$ and $\sigma_t \equiv \sigma(X_t)$. Cont & Mancini (2008) use TC to test for the presence of a diffusion component in asset prices. Gobbi & Mancini (2006, 2007) use TC to disentangle the co-jumps and the correlation between the diffusion parts of two semimartingales with Lévy type jumps. Aït-Sahalia & Jacod use TC to estimate the Blumenthal–Getoor index of J (Aït-Sahalia & Jacod, 2007) and to estimate the local volatility values σ_s^2 and σ_{s-}^2 to reach a test for the presence of jumps in a discretely observed semimartingale (Aït-Sahalia & Jacod, 2008).

5. Simulations

In this section we study the performance of our threshold estimator on finite samples. The aims of the section are on one hand to show that our technique is applicable and gives good results, on the other hand to make a comparison with the multipower variation performances. We do not pretend here to find an optimal threshold for each considered model. This is the topic of a further research.

We implement the threshold estimator within three different simulated models which are commonly used in finance: a jump diffusion process with jump part given by a compound Poisson process with Gaussian jump sizes; a model with FA jumps and a stochastic diffusion coefficient correlated with the Brownian motion driving the dynamics of X; and a model with constant volatility and an infinite activity (finite variation) Variance Gamma jump part. The models are described in detail below.

For each model, $N=5000$ trajectories have been generated, discretizing SDE (1), using the algorithms described in Cont & Tankov (2004), and taking $n=1000$ equally spaced observations X_{t_i} with lag $h=\frac{1}{n}$, corresponding to four observations per day, in a 7 hour open market, taken along 1 year ($T=1$). The adopted unit of measure is 1 year.

Remark: choice of the threshold. Our choice for the number n of observations to use and of the threshold $r(h)$ is assessed using simulations of the three models. This is done in the same spirit as for example Huang & Tauchen (2005) to show that, for finite samples, the log version of the bipower variation has a better performance than bipower variation; or in the same spirit as Aït-Sahalia & Jacod (2008) who use simulations to choose the threshold parameters α and ϖ as well as the parameters k, p and k_n involved in their estimators.

In fact, each model has its own optimal threshold, as is typical of all non-parametric estimators. Even for the multipower variation, the exponents r_i depend on the Blumenthal–Getoor index α of X (see below), which is unknown. However, our simulation study shows that $r(h)=h^{0.99}$ appears to be a good compromise along the three different models we present. In what follows we explain how we arrived at our choice. For $h\to 0$ we have that $(2h\ln\frac{1}{h})/h^{\beta}\to 0$ for all $\beta\in]0,1[$, so that the choice of a power of h seemed to be natural. The closer β is to 1, the closer is the speed of convergence to 0 of h^{β} to the speed of $2h\ln\frac{1}{h}$. However, for values of h in the range $[1/100{,}000; 1/252]$ in fact we have $2h\ln\frac{1}{h}>h^{\beta}$ for all values $\beta\in[0.9,1[$. This means that, in order to estimate IV, thresholds like $h^{0.9}$ or $h^{0.99}$ or $h^{0.999}$ exclude many observations which in fact are pertinent to estimate IV. If there are only rare jumps, then a higher $r(h)$ is better, since it includes more squared increments due only to the diffusion part. If, on the contrary, J has infinite activity then the variations $\Delta_i\tilde{J}_2$ are usually quite small, so the relative $(\Delta_i X)^2$ are included in $\widehat{\mathrm{IV}}_h$ and give a spurious estimate of IV if $r(h)$ is not properly small. In model 2 we even tried to use a variable threshold $r_{t_i}(h)$ linked to the value of $\sigma_{t_{i-1}}$ included in $\Delta_i X$. But in the end we concluded that $r(h)=h^{0.99}$ was the best choice.

Cont & Mancini (2008) consider a large number of different models and even there it is possible to take the same threshold for all the models.

We report in each top left panel of Figs 1, 2 and 3 the histogram of the 5000 values assumed in each model by the *normalized bias* term

$$\frac{\widehat{\mathrm{IV}}_h-\mathrm{IV}}{\sqrt{2h\widehat{\mathrm{IQ}}_h}} \tag{6}$$

versus the standard Gaussian density (continuous line).

For a comparison with the performance of the multipower variation estimators, define

$$V_{r_1,\ldots,r_k}(X):=\sum_{i=k}^{n}|\Delta_i X|^{r_1}|\Delta_{i-1}X|^{r_2}\ldots|\Delta_{i-k}X|^{r_k}. \tag{7}$$

From Woerner (2006) we know that if the Blumenthal–Getoor index[1] α of X is strictly less than 1 (which is the case for each of the three models used here), under both no leverage and given conditions on the drift part and on the volatility process σ (which are satisfied here in models 1 and 3), if $r_i > 0$ for all $i = 1..k$, $\max_i r_i < 1$ and $\sum_i r_i > \alpha/(2-\alpha)$, then as $h \to 0$,

$$\frac{h^{1-\sum_i r_i/2} V_{r_1,\dots,r_k}(X) - \prod_{i=1}^{k} \mu_{r_i} \int_0^T \sigma_u^{\sum_i r_i}\,\mathrm{d}u}{\sqrt{h}\sqrt{C_{r_1,\dots,r_k} \int_0^T \sigma_u^{2\sum_i r_i}\,\mathrm{d}u}} \xrightarrow{d} \mathcal{N}(0,1), \tag{8}$$

where $\mu_r = E[|Z|^r]$, Z is an $\mathcal{N}(0,1)$ random variable, and

$$C_{r_1,\dots,r_k} = \prod_{p=1}^{k} \mu_{2r_p} + 2\sum_{i=1}^{k-1} \prod_{p=1}^{i} \mu_{r_p} \prod_{p=k-i+1}^{k} \mu_{r_p} \prod_{p=1}^{k-i} \mu_{r_p + r_{p+i}} - (2k-1)\prod_{p=1}^{k} \mu_{r_p}^2.$$

A similar result in presence of leverage and when J is Lévy is obtained combining the outcomes in Barndorff-Nielsen *et al.* (2006a) and (2006b), allowing the application of the multipower variations even in model 2. Note that the integral in the denominator of (8) can be estimated using in turn the multipower variations.

In the light of the discussion at the beginning of section 3.3, we consider the two bipower variation estimators $V_{1,1}$ and $V_{\frac{1}{3},\frac{1}{3}}$, and the two tripower variation estimators $V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}}$ and $V_{0.99,0.02,0.99}$.

As for $V_{1,1}(X)$, Barndorff-Nielsen *et al.* (2006a) have shown that when X is a diffusion, then

$$\frac{\mu_1^{-2} V_{1,1}(X) - \int_0^T \sigma^2\,\mathrm{d}t}{\mu_1^{-2}\sqrt{C_{1,1}\mu_{\frac{4}{3}}^{-3} V_{\frac{4}{3},\frac{4}{3},\frac{4}{3}}(X)}} \tag{9}$$

is asymptotically Gaussian, while in presence of jumps, even if $\mu_1^{-2} V_{1,1}(X)$ is a consistent estimator of IV and $V_{\frac{4}{3},\frac{4}{3},\frac{4}{3}}(X)$ gives a consistent estimator of IQ, it has not be shown that such asymptotic normality holds. We study $V_{1,1}(X)$ anyway, since it is commonly used in practice. Huang & Tauchen (2005) have shown that, in absence of jumps, the modified log version $V_{1,1}$-Log$:= \log(\mu_1^{-2} V_{1,1}(X) \cdot n/(n-1))$ has a better performance than $\mu_1^{-2} V_{1,1}(X)$,[2] so in our comparisons we consider $V_{1,1}$-Log.

As for $V_{\frac{1}{3},\frac{1}{3}}(X)$, recall that it estimates $\int_0^T \sigma^{2/3}\,\mathrm{d}t$, and not directly $\int_0^T \sigma^2\,\mathrm{d}t$.

On the same 5000 generated paths of X we compute $V_{1,1}(X)$, $V_{\frac{1}{3},\frac{1}{3}}(X)$, $V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}}(X)$ and $V_{0.99,0.02,0.99}(X)$ and we plot in Figs 1–3 the histograms of the relative normalized biases, respectively,

$$\frac{\log(\mu_1^{-2} V_{1,1}(X) \cdot \frac{n}{n-1}) - \log(\int_0^T \sigma^2\,\mathrm{d}t)}{\mu_1^{-2}\sqrt{hC_{1,1}\max\left\{1, \frac{h^{-1}\mu_{\frac{4}{3}}^{-3} V_{\frac{4}{3},\frac{4}{3},\frac{4}{3}}(X)\cdot\frac{n}{n-2}}{\mu_1^{-4} V_{1,1}^2(X)\cdot\frac{n^2}{(n-1)^2}}\right\}}}, \tag{10}$$

[1]In Woerner (2006) the (generalized) Blumenthal–Getoor index of a semimartingale X is defined as $\alpha := \inf\{\delta > 0$: the process $\int_{s\in(0,t]}\int_{x\in\mathbb{R}} |x|^\delta \wedge 1\,\nu(\mathrm{d}x,\mathrm{d}s), t \geq 0$ is locally integrable$\}$, where ν is the compensator of the jump measure of X. This index measures how wild is the jump activity of X: the higher is α the more active is the jump component of X. In our framework, J is a Lévy process and α reduces to $\alpha := \inf\{\delta > 0 : \int_{x\in\mathbb{R}} |x|^\delta \wedge 1\,\nu(\mathrm{d}x) < \infty\}$.

[2]Using (9) and the delta-method we find that in absence of jumps

$$\frac{\log(\mu_1^{-2} V_{1,1}(X)) - \log(\int_0^T \sigma^2\,\mathrm{d}t)}{\mu_1^{-2}\sqrt{C_{1,1}\mu_{\frac{4}{3}}^{-3}\frac{V_{\frac{4}{3},\frac{4}{3},\frac{4}{3}}(X)}{\mu_1^{-4} V_{1,1}^2(X)}}} \xrightarrow{d} \mathcal{N}(0,1).$$

However, Huang & Tauchen (2005) find on simulations that the adjustments in (10) give better results.

$$\frac{h^{2/3}\mu_{\frac{1}{3}}^{-2} V_{\frac{1}{3},\frac{1}{3}}(X) - \int_0^T \sigma^{2/3}\,\mathrm{d}t}{\mu_{\frac{1}{3}}^{-2}\sqrt{hC_{\frac{1}{3},\frac{1}{3}} h^{1/3}\mu_{\frac{2}{3}}^{-2} V_{\frac{2}{3},\frac{2}{3}}(X)}}, \tag{11}$$

$$\frac{\mu_{\frac{2}{3}}^{-3} V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}}(X) - \int_0^T \sigma^2\,\mathrm{d}t}{\mu_{\frac{2}{3}}^{-3}\sqrt{hC_{\frac{2}{3},\frac{2}{3},\frac{2}{3}} h^{-1}\mu_{\frac{4}{5}}^{-5} V_{\frac{4}{5},\frac{4}{5},\frac{4}{5},\frac{4}{5},\frac{4}{5}}(X)}}, \tag{12}$$

$$\frac{\mu_{0.99}^{-2}\mu_{0.02}^{-1} V_{0.99,\,0.02,\,0.99}(X) - \int_0^T \sigma^2\,\mathrm{d}t}{\mu_{0.99}^{-2}\mu_{0.02}^{-1}\sqrt{hC_{0.99,\,0.02,\,0.99} h^{-1}\mu_{0.99}^{-4}\mu_{0.04}^{-1} V_{0.99,\,0.99,\,0.99,\,0.99,\,0.04}(X)}}. \tag{13}$$

For each model we show also the QQ-plots relative to the empirical distributions of (6), (10–13). All parameters are given on annual basis.

Model 1. We consider an FA jump diffusion model of kind

$$X_t = \sigma W_t + \sum_{j=1}^{N_t} Z_j, \quad t \in [0, T],$$

with Z_j i.i.d. with law $\mathcal{N}(0, \eta^2)$, where $\eta = 0.6$, $\sigma = 0.3$ and $\lambda = 5$, as in Aït-Sahalia (2004). Under this model Fig. 1 and Table 1 were generated.

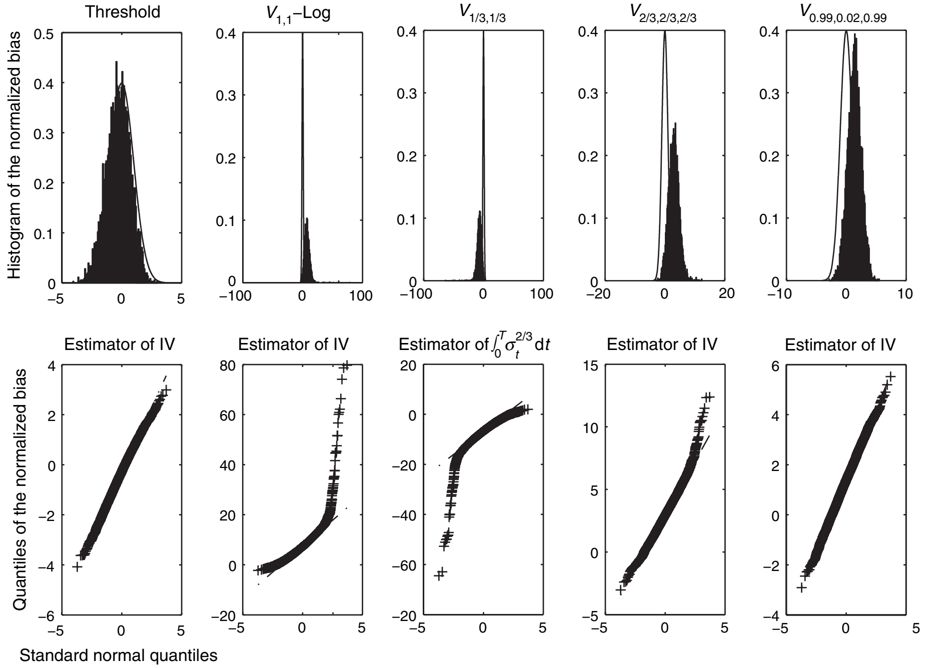

Fig. 1. Top, from left to right: histograms of 5000 values assumed by the normalized bias terms for the threshold estimator of IV as in (6), for the log bipower estimator $V_{1,1}$-Log of IV as in (10), for the bipower estimator $V_{1/3,\,1/3}$ of $\int_0^T \sigma_s^{2/3}\,\mathrm{d}s$ as in (11), for the tripower estimator $V_{2/3,\,2/3,\,2/3}$ of IV as in (12) and for the tripower estimator $V_{0.99,\,0.02,\,0.99}$ of IV as in (13), when X has constant volatility and compound Poisson jumps (Model 1). The continuous lines represent the density of the theoretical limit law $\mathcal{N}(0, 1)$. Bottom: corresponding QQ-plots. $n = 1000$, $T = 1$, $h = 1/n$, $r(h) = h^{0.99}$.

Table 1. *Statistics of the considered normalized biases under Model 1 and relative to the simulations in Fig. 1. Pct is the percentage of the 5000 realizations for which the normalized bias is in absolute value larger than 1.96 (asymptotically such a percentage has to be 0.05). Mean and SD are the mean and the standard deviation of the 5000 values assumed by each normalized bias term (6), (10), (11), (12) and (13) (asymptotically such mean and SD have to be 0 and 1)*

	Threshold	$V_{1,1}$-Log	$V_{1/3,\,1/3}$	$V_{2/3,\,2/3,\,2/3}$	$V_{0.99,\,0.02,\,0.99}$
Pct	0.0644	0.9276	0.3004	0.7156	0.9294
Mean	−0.3097	−7.1821	1.3581	2.9663	7.7804
SD	1.0146	4.5698	1.1435	1.7672	5.2413

We remark that the smaller is η the better is the performance of $V_{1,1}$-Log. The difference between the threshold and the bipower is particularly evident in presence of big jumps, when h is not very small.

Model 2. We now consider a process with FA jump part as in model 1 and with a stochastic diffusion coefficient correlated with the Brownian motion driving X:

$$X_t = \ln(S_t),$$

where

$$\frac{dS_t}{S_{t-}} = (\mu + \sigma_t^2/2)\,dt + \sigma_t\,dW_t^{(1)} + d\bar{J}_t, \quad \bar{J}_t = \sum_{j=1}^{N_t} \bar{Z}_j, \quad \ln(1+\bar{Z}_j) \sim \mathcal{N}(m_G, \eta^2),$$

$$\sigma_t = e^{H_t}, \quad dH_t = -k(H_t - \bar{H})\,dt + v\,dW_t^{(2)}, \quad d < W^{(1)}, W^{(2)} >_t = \rho\,dt.$$

Note that

$$dX_t = \mu\,dt + \sigma_t\,dW_t^{(1)} + dJ_t,$$

where

$$J_t = \sum_{j=1}^{N_t} Z_j, \quad Z_j \sim \mathcal{N}(m_G, \eta^2).$$

We chose $\mu = 0$, $\lambda = 5$ and a constant negative correlation coefficient $\rho = -0.7$; then we took $H_0 \equiv \ln(0.1)$, $k = 3$, $\bar{H} = \ln(0.2)$, $v = 0.6$ so that a path of σ within $[0, T]$ varies most between 10% and 50%. We remark that with the choice of parameters as for SVJ1F model in Huang & Tauchen (2005), Fig. 2, the range of the realized path of σ is mainly $[0.53, 1.87]$ (percent values) on a daily basis (the volatility factor varies most within $[-4, 8]$, then $\sigma_t = e^{\beta_1 v_t}$ varies most within $[0.53, 1.87]$), corresponding to a range $[9.6, 43]$ (percent values) on annual basis. Moreover $m_G = 0$ and $\eta = 0.6$ give relative amplitudes of the jumps of S, in absolute value, most between 0.01 and 0.60. Under this model Fig. 2 and Table 2 were generated.

Model 3. The underlying model for X is given by a diffusion plus a Variance Gamma (VG) jump component. The VG process is a pure jump process with infinite activity and finite variation and is given by $cG_t + \eta W_{G_t}$, the composition of a Brownian motion with drift and an independent Gamma process G. G is such that for each h the r.v. G_h at time h has law gamma: $(G_h)(P) = \text{GAMMA}(h/b, b)$, where $b = \text{var}(G_1)$. c and η are constants. We add an independent diffusion component σB_t, where σ is constant and B is a Brownian motion, so that

$$X_t = \sigma B_t + cG_t + \eta W_{G_t}.$$

$b = 0.23$, $c = -0.2$ and $\eta = 0.2$ are as in Madan (2001); $\sigma = 0.3$ is chosen as in model 1. Under this model, Fig. 3 and Table 3 were generated.

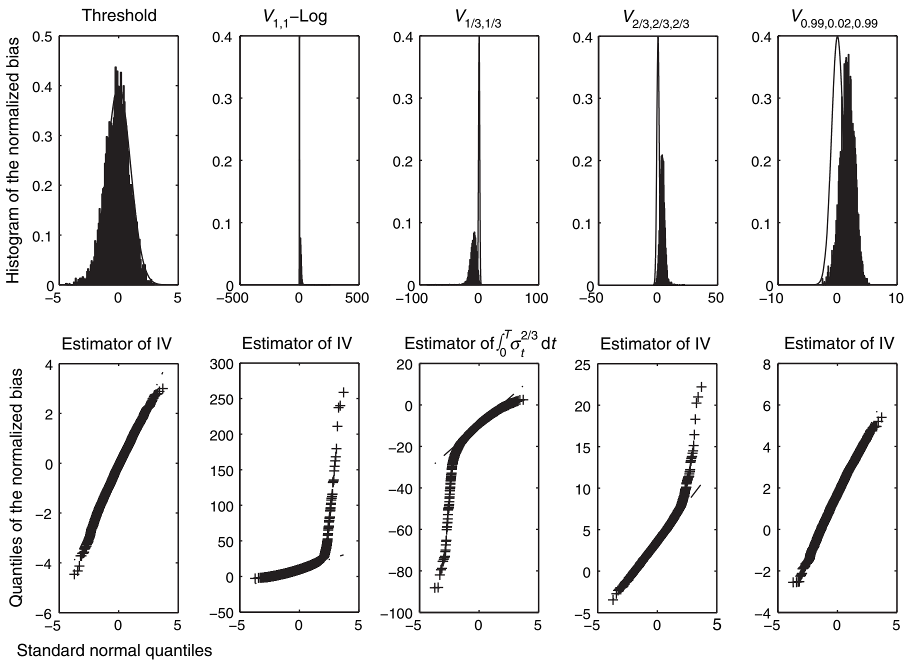

Fig. 2. Top, from left to right: histograms of 5000 values assumed by the normalized bias terms for the threshold estimator of IV as in (6), for the log bipower estimator $V_{1,\,1}$-Log of IV as in (10), for the bipower estimator $V_{1/3,\,1/3}$ of $\int_0^T \sigma_s^{2/3}\,\mathrm{d}s$ as in (11), for the tripower estimator $V_{2/3,\,2/3,\,2/3}$ of IV as in (12) and for the tripower estimator $V_{0.99,\,0.02,\,0.99}$ of IV as in (13), when X has stochastic volatility, negatively correlated with the Brownian motion in the dynamics of X, plus compound Poisson jumps (Model 2). The continuous lines represent the density of the theoretical limit law $\mathcal{N}(0,1)$. Bottom: corresponding QQ-plots. $n=1000$, $T=1$, $h=1/n$, $r(h)=h^{0.99}$.

Table 2. *Statistics of the considered normalized biases under Model 2 and relative to the simulations in Fig. 2. See Table 1 for explanation*

	Threshold	$V_{1,\,1}$-Log	$V_{1/3,\,1/3}$	$V_{2/3,\,2/3,\,2/3}$	$V_{0.99,\,0.02,\,0.99}$
Pct	0.0558	0.9586	0.4122	0.8056	0.9624
Mean	−0.1260	−10.1758	1.6686	3.6479	11.3959
SD	1.0235	6.9555	1.1895	2.0914	12.1050

Figs 4, 5 and 6 show the histograms and QQ-plots relative to (6), (10–13) for models 1, 2, 3 when we use smaller lag h and time horizon of 1 day: $n=288$ and $h=1/(252\times 288)$, corresponding to 288 observations in 1 day. The threshold is still $r(h)=h^{0.99}$. Corresponding summary statistics are reported in Tables 4, 5 and 6, respectively.

Remarks. The case $n=1000$, $h=1/1000$. The figures and the QQ-plots show that, for $n=1000$ and $h=1/1000$, our $\widehat{\mathrm{IV}}_h$ is much less biased and has much better standard error and pct (as defined in the caption of Table 1, pct is the percentage of the 5000 realizations of the normalized bias which are larger than 1.96 in absolute value) than the MPVs.

The best estimator among the multipower variations is $V_{\frac{1}{3},\,\frac{1}{3}}$, but note that it in fact gives an estimate of $\int_0^T \sigma_s^{2/3}\,\mathrm{d}s$, and when σ is not constant (which is the most plausible case

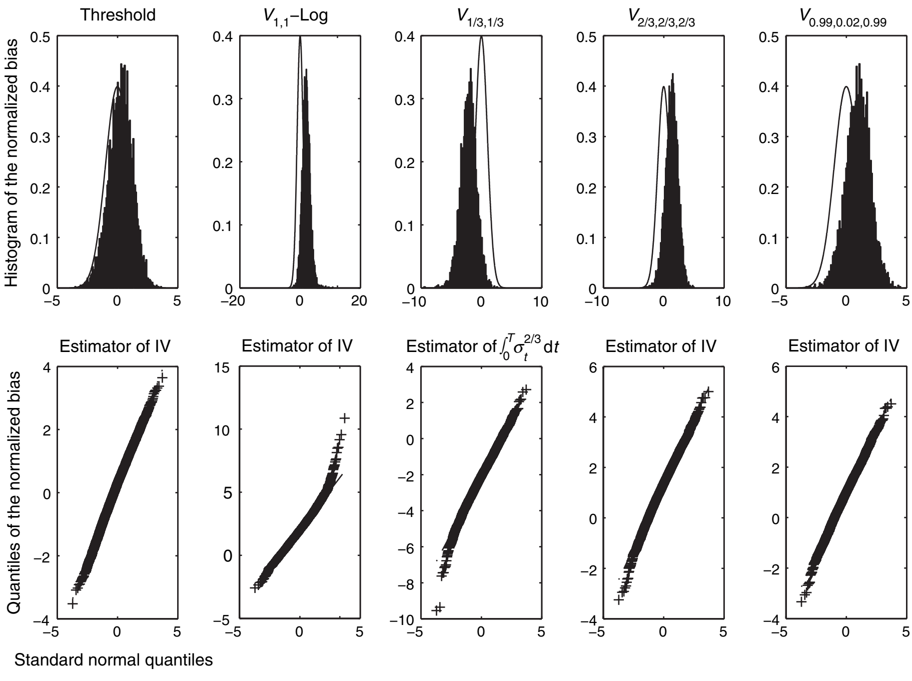

Fig. 3. Top, from left to right: histograms of 5000 values assumed by the normalized bias terms for the threshold estimator of IV as in (6), for the log bipower estimator $V_{1,\,1}$-Log of IV as in (10), for the bipower estimator $V_{1/3,\,1/3}$ of $\int_0^T \sigma_s^{2/3}\,\mathrm{d}s$ as in (11), for the tripower estimator $V_{2/3,\,2/3,\,2/3}$ of IV as in (12) and for the tripower estimator $V_{0.99,\,0.02,\,0.99}$ of IV as in (13), when X has constant volatility plus a Variance Gamma jump part (Model 3). The continuous lines represent the density of the theoretical limit law $\mathcal{N}(0,1)$. Bottom: corresponding QQ-plots. $n=1000$, $T=1$, $h=1/n$, $r(h)=h^{0.99}$.

Table 3. *Statistics of the considered normalized biases under Model 3 and relative to the simulations in Fig. 3. See Table 1 for explanation*

	Threshold	$V_{1,\,1}$-Log	$V_{1/3,\,1/3}$	$V_{2/3,\,2/3,\,2/3}$	$V_{0.99,\,0.02,\,0.99}$
Pct	0.0536	0.5402	0.1606	0.2760	0.5126
Mean	0.2570	−2.1264	0.9677	1.3172	2.0289
SD	0.9850	1.3110	1.0143	1.0433	1.3096

in a realistic situation) to convert this to an estimate of IV some new bias would be introduced.

Compared with $\widehat{\mathrm{IV}}_h$ the tails of $V_{0.99,\,0.02,\,0.99}$ are good, but the estimator is biased in all the three models. The same is true even for $V_{\frac{1}{3},\,\frac{1}{3}}$ and $V_{\frac{2}{3},\,\frac{2}{3},\,\frac{2}{3}}$ in model 3.

The slight gain in efficiency passing from $V_{\frac{2}{3},\,\frac{2}{3},\,\frac{2}{3}}$ to $V_{0.99,\,0.02,\,0.99}$ is not visible for $n=1000$: the pct of $V_{0.99,\,0.02,\,0.99}$ is in fact worse.

$V_{\frac{2}{3},\,\frac{2}{3},\,\frac{2}{3}}$ is much worse than $\widehat{\mathrm{IV}}_h$.

$V_{1,\,1}$-Log is not reliable in presence of jumps. It is better in model 3, since in fact the Variance Gamma process has small jumps.

Note that for model 1 we do not reach a pct of 5% for $\widehat{\mathrm{IV}}_h$, as we would expect. This is because for this model $r(h)=h^{0.99}$ is in fact non-optimal. A threshold $r(h)=h^{0.9}$ gives a better result (in simulations that we do not present here). However, here we prefer to have the same threshold for all the models, since when we apply the estimator to market data we do not know which model the data-generating process follows.

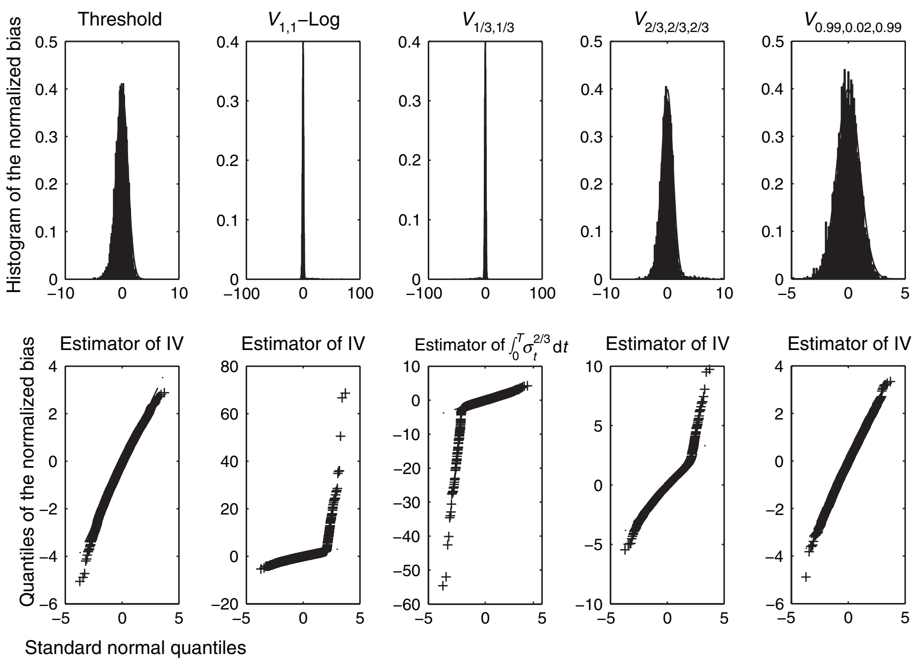

Fig. 4. Top: histograms of 5000 values assumed by the normalized bias terms (6), (10), (11), (12) and (13), under Model 1. Bottom: corresponding QQ-plots. $n = 288$, $T = nh$, $h = 1/(288 \times 252)$, $r(h) = h^{0.99}$.

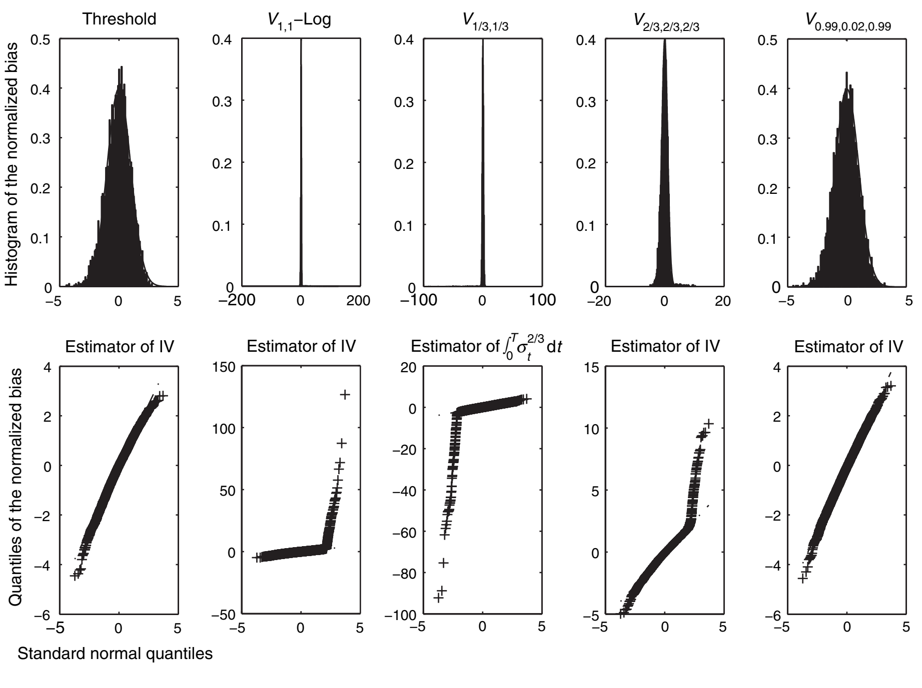

Fig. 5. Top: histograms of 5000 values assumed by the normalized bias terms (6), (10), (11), (12) and (13), under Model 2. Bottom: corresponding QQ-plots. $n = 288$, $T = nh$, $h = 1/(288 \times 252)$, $r(h) = h^{0.99}$.

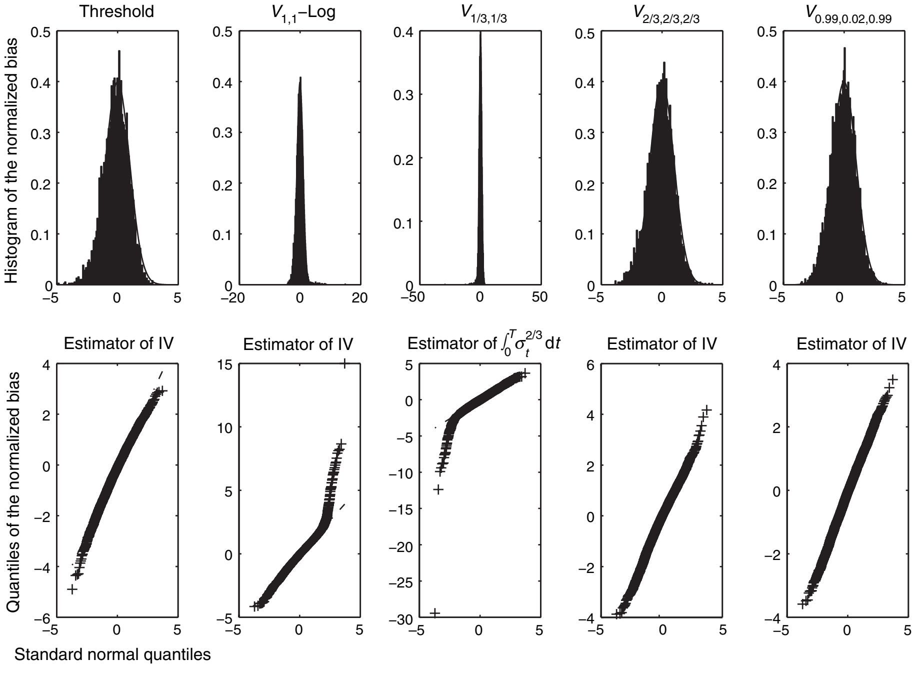

Fig. 6. Top: histograms of 5000 values assumed by the normalized bias terms (6), (10), (11), (12) and (13), under Model 3. Bottom: corresponding QQ-plots. $n=288$, $T=nh$, $h=1/(288\times 252)$, $r(h)=h^{0.99}$.

Table 4. *Statistics of the considered normalized biases under Model 1 and relative to the simulations in Fig. 4*

	Threshold	$V_{1,1}$-Log	$V_{1/3,1/3}$	$V_{2/3,2/3,2/3}$	$V_{0.99,0.02,0.99}$
Pct	0.0526	0.0672	0.0538	0.0730	0.0742
Mean	−0.1297	−0.2686	−0.0735	−0.0800	0.1382
SD	1.0165	2.7327	1.0064	1.1760	2.8122

Table 5. *Statistics of the considered normalized biases under Model 2 and relative to the simulations in Fig. 5*

	Threshold	$V_{1,1}$-Log	$V_{1/3,1/3}$	$V_{2/3,2/3,2/3}$	$V_{0.99,0.02,0.99}$
Pct	0.0510	0.0712	0.0532	0.0720	0.0764
Mean	−0.1000	0.0027	−0.0968	−0.1174	0.1147
SD	1.0074	1.5393	1.0136	1.1570	2.8722

Table 6. *Statistics of the considered normalized biases under Model 3 and relative to the simulations in Fig. 6*

	Threshold	$V_{1,1}$-Log	$V_{1/3,1/3}$	$V_{2/3,2/3,2/3}$	$V_{0.99,0.02,0.99}$
Pct	0.0574	0.0734	0.0556	0.0714	0.0840
Mean	−0.1563	−0.0869	−0.0986	−0.1211	−0.0541
SD	1.0293	1.2609	1.0114	1.0566	1.2324

The case $n=288$, $h=1/(252\times 288)$. In this case, the performance of the threshold estimator is essentially as when $h=1/1000$, while the MPVs improve very much. However, $\widehat{\mathrm{IV}}_h$ is still globally better to estimate IV.

6. Conclusions

In this paper we devise a technique, based on discrete observations, for identifying the time instants of significant jumps for a process driven by diffusions and jumps. The technique makes use of a suitably defined threshold. When J has FA, we give a non-parametric estimate of the jump times and sizes, while when J has an IA jump Lévy component we can identify the instants when jumps are larger than the threshold. As a consequence we provide a consistent estimate of $\mathrm{IV}=\int_0^T \sigma_t^2\,\mathrm{d}t$, extending previous results (Mancini, 2001, 2004) with very mild assumptions on a and σ and allowing for infinite jump activity. When J has FA we also prove central limit results for the family $\{\widehat{\mathrm{IV}}_h\}_h$ and for the jump size estimates.

Compared with power variations, multipower variations or kernel estimators, the threshold estimator is efficient. The inefficiency of the multipower variations is evident for values of $h=1/1000$ up to $h=1/20{,}000$ and for small volatility values (Cont & Mancini, 2008). In the FA case, the threshold method is a more effective way to identify each interval $]t_{j-1},t_j]$ where J jumped. In particular, it allows the extension of kernel estimators used in diffusion frameworks to processes driven by diffusions and jumps, provided we eliminate the jumps (Mancini & Renò, 2006). $\widehat{\mathrm{IV}}_h$ is consistent both in the FA and in the IA cases. The threshold technique works even when the observations are not equally spaced and also when the threshold is time varying, whatever be the dynamics for σ. The approach presented here allows to prove a central limit theorem for $\widehat{\mathrm{IV}}_h$, for any càdlàg process σ (for the case where J has IA, see Cont & Mancini, 2008).

The good performance of our estimator on finite samples with both ‘large’ and small h is shown within three different simulated models which are commonly used in practice. The choice of the threshold is assessed on simulations.

A comparison with the multipower variation estimators shows that $\widehat{\mathrm{IV}}_h$ gives more reliable results, especially when the frequency of the observations is not so high.

Acknowledgements

I am sincerely grateful to Rama Cont, Jean Jacod and two anonymous referees for the important comments on this work. A special thank to Roberto Renò for the many and useful discussions and for the constant encouragement. I also thank Giampiero Maria Gallo, Maria Elvira Mancino and PierLuigi Zezza for the help and advises. I thank Sergio Vessella and Marcello Galeotti who supported this research by MIUR grant number 2002013279 and Progetto Strategico.

References

Aït-Sahalia, Y. (2004). Disentangling volatility from jumps. *J. Financ. Econ.* **74**, 487–528.

Aït-Sahalia, Y. & Jacod, J. (2007). Estimating the degree of activity of jumps in high frequency financial data. Working paper, Princeton University and Université de Paris-6, available on http://www.princeton.edu/yacine/research.htm

Aït-Sahalia, Y. & Jacod, J. (2008). Testing for jumps in a discretely observed process (forthcoming in *Ann. Statist*).

Andersen, T. G., Benzoni, L. & Lund, J. (2002). An empirical investigation of continuous-time models for equity returns. *J. Finance* **57**, 1239–1284.

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2005). Parametric and non-parametric volatility measurement. In *Handbook of financial econometrics* (eds Y. Aït-Sahalia & L. P. Hansen). North Holland, Amsterdam.

Bandi, F. M. & Nguyen, T. H. (2003). On the functional estimation of jump-diffusion models. *J. Econom.* **116**, 293–328.

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

Barndorff-Nielsen, O. E. & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *J. Financ. Econom.* **4**, 1–30.

Barndorff-Nielsen, O. E. & Shephard, N. (2007). Variation, jumps and high frequency data in financial econometrics. In *Advances in economics and econometrics. Theory and applications. Ninth world congress* (eds R. Blundell, P. Torsten & K. W. Newey), 328–372. Econometric Society Monographs, Cambridge University Press.

Barndorff-Nielsen, O. E., Gravensen, S. E., Jacod, J., Podolskij, M. & Shephard, N. (2006a). A central limit theorem for realised power and bipower variation of continuous semimartingales. In *From stochastic analysis to mathematical finance. Festschrift for Albert Shiryaev* (eds Y. Kabanov, R. Lipster & J. Stoyanov), 33–68. Springer, Berlin.

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

Berman, S. M. (1965). Sign-invariant random variables and stochastic processes with sign invariant increments. *Trans. Amer. Math. Soc.* **119**, 216–243.

Chung, K. L. (1974). *A course in probability theory*. Academic Press, New York.

Cont, R. & Mancini, C. (2008). Non-parametric tests for analyzing the fine structure of price fluctuations. Available on http://www.ssrn.com

Cont, R. & Tankov, P. (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC, Boca Raton.

Cont, R., Tankov, P. & Voltchkova, E. (2007). Hedging with options in presence of jumps. In *Stochastic analysis and applications, the Abel symposium 2005* (eds F. Benth, G. Di Nunno, T. Lindstrøm, B. Øksendal & T. Zhang), 197–218. Springer, Berlin.

Das, S. (2002). The surprise element: jumps in interest rates. *J. Econom.* **106**, 27–65.

Eraker, B., Johannes, M. & Polson, N. (2003). The impact of jumps in equity index volatility and returns. *J. Finan.* **58**, 1269–1300.

Gobbi, F. & Mancini, C. (2006). Identifying the diffusion covariation and the co-jumps given discrete observations. arXiv:math/0610621v1 [math.PR], submitted.

Gobbi, F. & Mancini, C. (2007). Diffusion covariation and co-jumps in bidimensional asset price processes with stochastic volatility and infinite activity Lévy jumps. arXiv:0705.1268v1 [math.PR].

Huang, X. & Tauchen, G. (2005). The relative contribution of jumps to total price variance. *J. Finan. Econom.* **3**, 456–499.

Ikeda, N. & Watanabe, S. (1981). *Stochastic differential equations and diffusion processes*. North Holland, Amsterdam.

Jacod, J. (2008). Asymptotic properties of realized power variations and associated functionals of semimartingales. *Stoch. Proc. Appl.* **118**, 517–559.

Jacod, J. & Protter, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26**, 267–307.

Johannes, M. (2004). The statistical and economic role of jumps in continuous-time interest rate models. *J. Finance* **59**, 227–260.

Karatzas, I. & Shreve, S. E. (1999). *Brownian motion and stochastic calculus*. Springer, New York.

Madan, D. B. (2001). Purely discontinuous asset price processes. In *Handbooks in mathematical finance: option pricing, interest rates and risk management* (eds J. Cvitanic, E. Jouini & M. Musiela), 105–153. Cambridge University Press, Cambridge.

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

Mancini, C. (2004). Estimation of the characteristics of the jumps of a general Poisson-diffusion model. *Scand. Actuar. J.* **1**, 42–52.

Mancini, C. & Renò, R. (2006). Threshold estimation of jump-diffusion models and interest rate modeling. Working paper of the University of Siena.

Métivier, M. (1982). *Semimartingales: a course on stochastic processes*. De Gruyter, New York.

Protter, P. (1990). *Stochastic integration and differential equations*. Springer-Verlag, Berlin.

Revuz, D. & Yor, M. (2001). *Continuous martingales and Brownian motion*. Springer, New York.

Sato, K. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge University Press, Cambridge.

Woerner, J. (2006). Power and multipower variation: inference for high frequency data. In *Stochastic finance* (eds A. N. Shiryaev, M. R. Grossinho, P. Oliveira, & M. Esquivel), 343–364. Springer, New York.

Received August 2006, in final form June 2008

Cecilia Mancini, Università di Firenze, Dipartimento di matematica per le decisioni, via C. Lombroso, 6/17, 50134 Florence, Italy.
E-mail: cecilia.mancini@dmd.unifi.it

Appendix: proofs

Proof of theorem 1. For the proof we need the following preliminary remarks.

(i) The Paul Lévy law for the modulus of continuity of Brownian motion's paths (Karatzas & Shreve, 1999, p. 114, theorem 9.25) implies that

$$\text{a.s.} \quad \lim_{h\to 0} \sup_{i\in\{1,\ldots,n\}} \frac{|\Delta_i W|}{\sqrt{2h\log\frac{1}{h}}} \leq 1.$$

(ii) The stochastic integral $\sigma\cdot W$ is a time changed Brownian motion (Revuz & Yor, 2001, theorems 1.9 and 1.10): $\Delta_i(\sigma\cdot W)=B_{\mathrm{IV}_{t_i}}-B_{\mathrm{IV}_{t_{i-1}}}$, where B is a Brownian motion.

(iii) As a consequence, under assumptions 1 and 2 of theorem 1, a.s. for small h

$$\sup_{i\in\{1,\ldots,n\}} \frac{|\int_{t_{i-1}}^{t_i} a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s|}{\sqrt{2h\log\frac{1}{h}}} \leq \Lambda(\omega), \tag{14}$$

where $\Lambda(\omega)=C(\omega)+\sqrt{M(\omega)}+1$ is a finite r.v. In fact, a.s.

$$\sup_{i\in\{1,\ldots,n\}} \frac{|\int_{t_{i-1}}^{t_i} a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s|}{\sqrt{2h\log\frac{1}{h}}} \leq \sup_{i\in\{1,\ldots,n\}} \frac{|\int_{t_{i-1}}^{t_i} a_s\,\mathrm{d}s|}{\sqrt{2h\log\frac{1}{h}}} + \sup_{i\in\{1,\ldots,n\}} \frac{|\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s|}{\sqrt{2h\log\frac{1}{h}}}$$

$$\leq C(\omega)+\sup_{i\in\{1,\ldots,n\}} \frac{|B_{\mathrm{IV}_{t_i}}-B_{\mathrm{IV}_{t_{i-1}}}|}{\sqrt{2\Delta_i\mathrm{IV}\log\frac{1}{\Delta_i IV}}} \sup_{i\in\{1,\ldots,n\}} \frac{\sqrt{2\Delta_i\mathrm{IV}\log\frac{1}{\Delta_i IV}}}{\sqrt{2Mh\log\frac{1}{Mh}}} \sup_{i\in\{1,\ldots,n\}} \frac{\sqrt{2Mh\log\frac{1}{Mh}}}{\sqrt{2h\log\frac{1}{h}}},$$

and by Karatzas & Shreve (1999, theorem 9.25), and the monotonicity of the function $x\ln(1/x)$, it follows that as $h\to 0$, the lim sup of the right-hand side is bounded by $C(\omega)+\sqrt{M(\omega)}$. Thus, for sufficiently small h (14) holds.

Now we come to prove the theorem. First, we show that a.s., for small h, we have $\forall i, I_{\{\Delta_i N=0\}} \leq I_{\{(\Delta_i X)^2\leq r(h)\}}$. Then we see that a.s., for small h, it holds also that $\forall i, I_{\{\Delta_i N=0\}} \geq I_{\{(\Delta_i X)^2\leq r(h)\}}$, and that concludes our proof.

(1) For each ω set $J_{0,h}=\{i\in\{1,\ldots,n\}:\Delta_i N=0\}$: to show that a.s., for small h, $I_{\{\Delta_i N=0\}}\leq I_{\{(\Delta_i X)^2\leq r(h)\}}$ it is sufficient to prove that a.s., for small h, $\sup_{J_{0,h}}(\Delta_i X)^2\leq r(h)$. To evaluate the $\sup_{J_{0,h}}(\Delta_i X)^2$, remark that a.s.

$$\sup_{i\in J_{0,h}} \frac{(\Delta_i X)^2}{r(h)} = \sup_{J_{0,h}} \left(\frac{|\int_{t_{i-1}}^{t_i} a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s|}{\sqrt{2h\log\frac{1}{h}}}\right)^2 \cdot \frac{2h\log\frac{1}{h}}{r(h)} \leq \Lambda^2\frac{2h\log\frac{1}{h}}{r(h)} \to 0, \quad \text{as } h\to 0.$$

In particular, for small h, $\sup_{i\in J_{0,h}}(\Delta_i X)^2/r(h)\leq 1$, as we need.

Note that if $N_T=0$ then all indexes i fall into the class $J_{0,h}$ of point (1), and so, for small h, we never have $(\Delta_i X)^2>r(h)$.

(2) Now we establish the other inequality. For any ω set $J_{1,h}=\{i\in\{1,\dots,n\}:\Delta_i N\neq 0\}$. In order to prove that a.s., for small h, $\forall i$, $I_{\{\Delta_i N=0\}}\geq I_{\{(\Delta_i X)^2\leq r(h)\}}$, it is sufficient to show that a.s., for small h, $\inf_{i\in J_{1,h}}(\Delta_i X)^2>r(h)$. In order to evaluate $\inf_{i\in J_{1,h}}(\Delta_i X)^2/r(h)$ remark that

$$\forall i\in J_{1,h},\ \frac{(\Delta_i X)^2}{r(h)}=\frac{\left(\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\Delta_i\sigma\cdot W\right)^2}{2h\log\frac{1}{h}}\frac{2h\log\frac{1}{h}}{r(h)}$$
$$+2\frac{\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\Delta_i\sigma\cdot W}{\sqrt{r(h)}}\frac{\sum_{\ell=1}^{\Delta_i N}\gamma_\ell}{\sqrt{r(h)}}+\frac{(\sum_{\ell=1}^{\Delta_i N}\gamma_\ell)^2}{r(h)}.$$

The first term tends a.s. to zero uniformly with respect to i. Since for small h we have that $\Delta_i N\leq 1$ for each i, then the other terms become

$$\frac{\gamma_{\tau(i)}}{\sqrt{r(h)}}\left[2\frac{\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\Delta_i(\sigma\cdot W)}{\sqrt{r(h)}}+\frac{\gamma_{\tau(i)}}{\sqrt{r(h)}}\right].$$

The contribution of the first term within brackets tends a.s. to zero uniformly in i. Note that the assumption on J guarantees that $P\{\underline{\gamma}=0\}=0$, thus a.s.

$$\lim_h\inf_{i\in J_{1,h}}\frac{(\Delta_i X)^2}{r(h)}=\lim_h\frac{\gamma_{\tau(i)}^2}{r(h)}\geq\lim_h\frac{\underline{\gamma}^2}{r(h)}=+\infty.$$

Remark for the case of not equally spaced observations. When the lag $\Delta t_i:=t_i-t_{i-1}$ between the observations $\{x_0,X_{t_1},\dots,X_{t_{n-1}},X_{t_n}\}$ is not constant, then defining $h:=\max_i\Delta t_i$, theorem 1 and corollary 2 are still valid. In fact all the fundamental tools of the proof of theorem 1 hold:

$$\lim_{h\to 0}\sup_{i\in\{1,\dots,n\}}\frac{|\Delta_i W|}{\sqrt{2h\log\frac{1}{h}}}\leq\lim_{h\to 0}\sup_{i\in\{1,\dots,n\}}\frac{|\Delta_i W|}{\sqrt{2\Delta t_i\log\frac{1}{\Delta t_i}}}\leq 1,$$

by the monotonicity of $x\ln\frac{1}{x}$. Moreover, using the Brownian motion time change as in the previous proof, it still holds that a.s., for small h,

$$\sup_{i\in\{1,\dots,n\}}\frac{|\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s|}{\sqrt{2h\log\frac{1}{h}}}\leq\Lambda(\omega),$$

since, for all i, $\Delta_i\mathrm{IV}<\Delta t_i\cdot M(\omega)\leq hM(\omega)$. As a consequence, the proof of the previous points (1) and (2) proceed analogously.

Alternatively, and equivalently, we can directly compare each $(\Delta_i X)^2$ with the relative $r(\Delta t_i)$. In fact, under the assumptions stated in corollary 1, it still holds that a.s., for small $h:=\max_i\Delta t_i$,

$$\sup_{i\in\{1,\dots,n\}}\frac{|\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s|}{\sqrt{2\Delta t_i\log\frac{1}{\Delta t_i}}}\leq\Lambda(\omega),$$

and the proofs of points (1) and (2) follow.

Proof of corollary 2. Since a.s. for small h we have $I_{\{\Delta_i N=0\}}=I_{\{(\Delta_i X)^2\le r(h)\}}$, for all $i=1..n$, then

$$\operatorname*{Plim}_{h\to 0}\sum_{i=1}^{n}(\Delta_i X)^2 I_{\{(\Delta_i X)^2\le r(h)\}}=\operatorname*{Plim}_{h\to 0}\sum_{i=1}^{n}(\Delta_i X)^2 I_{\{\Delta_i N=0\}}$$
$$=\operatorname*{Plim}_{h\to 0}\sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s\right)^2-\operatorname*{Plim}_{h\to 0}\sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s\right)^2 I_{\{\Delta_i N\neq 0\}},$$

which coincides with $\int_0^T\sigma_u^2\,\mathrm{d}u$, since a.s. $\sum_{i=1}^n(\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s)^2 I_{\{\Delta_i N\neq 0\}}\le N_T\sup_i(\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s)^2\to 0$, as $h\to 0$.

Theorem 5 (*Power variation* estimator: theorem 2.2 in Barndorff-Nielsen *et al.*, 2006a, case $r=4, s=0$)
If $\mathrm{d}X_s=a_s\,\mathrm{d}s+\sigma_s\,\mathrm{d}W_s$, *where* a *is predictable and locally bounded and* σ *is càdlàg, then for* $h\to 0$,

$$\frac{1}{3h}\sum_{i=1}^{n}(\Delta_i X)^4\xrightarrow{P}\int_0^T\sigma_t^4\,\mathrm{d}t.$$

Proof of proposition 1. By theorem 1, as $h\to 0$

$$\operatorname*{Plim}_{h\to 0}\frac{1}{3h}\sum_{i=1}^{n}(\Delta_i X)^4 I_{\{(\Delta_i X)^2\le r(h)\}}=\operatorname*{Plim}_{h\to 0}\frac{1}{3h}\sum_{i=1}^{n}(\Delta_i X)^4 I_{\{\Delta_i N=0\}}.$$

The latter coincides with

$$\operatorname*{Plim}_{h\to 0}\frac{1}{3h}\sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s\right)^4, \tag{15}$$

since

$$\operatorname*{Plim}_{h\to 0}\frac{1}{3h}\sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s\right)^4 I_{\{\Delta_i N\neq 0\}}\le\operatorname*{Plim}_{h\to 0}\Lambda^4 N_T\frac{(h\ln\frac{1}{h})^2}{3h}=0. \tag{16}$$

Finally note that $\int_0^t a_s\,\mathrm{d}t=\int_0^t a_{s-}\,\mathrm{d}t$ and that when a is càdlàg then a_- is predictable and locally bounded. Hence we can apply theorem 2.2 in Barndorff-Nielsen *et al.* (2006a) and conclude that (15) coincides with $\int_0^T\sigma_t^4\,\mathrm{d}t$.

Theorem 6 (Theorem 1 in Barndorff-Nielsen & Shephard, 2007)
If $\mathrm{d}X=a_s\,\mathrm{d}s+\sigma_s\,\mathrm{d}W_s$, *where* a *and* σ *are càdlàg processes, then, as* $h\to 0$,

$$\frac{[X^{(h)}]_T-[X]_T}{\sqrt{h}}\to\sqrt{2}\int_0^T\sigma_u^2\,\mathrm{d}B_u,$$

stably in law, where B *is a Brownian motion independent of* X *(recall from the notation that* $[X^{(h)}]_T=\sum_{i=1}^n(\Delta_i X)^2$*).*

Proof of theorem 2. Denote by X_0 the continuous process given by $X_{0t}=\int_0^t a_s\,\mathrm{d}s+\int_0^t\sigma_s\,\mathrm{d}W_s$, for all $t\in[0,T]$. To prove theorem 6, Barndorff-Nielsen & Shephard (2007) use some Itô algebra to show that

$$\frac{[X_0^{(h)}]_T-[X_0]_T}{\sqrt{h}}=2\frac{\int_0^{h[T/h]}X_{0u}-X_{0\,h[u/h]}\,\mathrm{d}X_{0u}}{\sqrt{h}},$$

and then they use that $2(\int_0^T X_{0u} - X_{0h[u/h]}\,\mathrm{d}X_{0u})/\sqrt{h}$ stably converges towards $\sqrt{2}\int_0^T \sigma_t^2\,\mathrm{d}B_t$ (Jacod & Protter, 1998, theorem 5.5), where B is a Brownian motion independent of the space where X_0 is defined.

However, the last mentioned result holds whatever be the space $(\Omega, (\mathcal{F}_t)_t, \mathcal{F})$ where X_0 is defined, as soon as on such a space X_0 is a continuous Itô semimartingale (i.e. with absolutely continuous local characteristics: $X_0 = A + M$, where $A_t = \int_0^t a_s\,\mathrm{d}s$ and $<M, M>_t = \int_0^t \sigma_s^2\,\mathrm{d}s$) having a.s. $\int_0^T a_s^2\,\mathrm{d}s < \infty$, and $\int_0^T \sigma_s^4\,\mathrm{d}s < \infty$. Since this is the case under the assumptions of our theorem 2, then we still have the stable convergence

$$\frac{[X_0^{(h)}]_T - [X_0]_T}{\sqrt{h}} \xrightarrow{st} \sqrt{2}\int_0^T \sigma_t^2\,\mathrm{d}B_t,$$

even in our space where also a jump process is defined; and B is a Brownian motion independent on our whole space $(\Omega, (\mathcal{F}_t)_t, \mathcal{F}, P)$.

Thus in particular we have

$$\left(\frac{[X_0^{(h)}]_T - [X_0]_T}{\sqrt{h}}, \int_0^T \sigma_u^4\,\mathrm{d}u\right) \xrightarrow{d} \left(\sqrt{2}\int_0^T \sigma_t^2\,\mathrm{d}B_t, \int_0^T \sigma_u^4\,\mathrm{d}u\right)$$

and so

$$\frac{[X_0^{(h)}]_T - [X_0]_T}{\sqrt{2h\int_0^T \sigma_u^4\,\mathrm{d}u}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Now, using theorem 1,

$$\frac{\widehat{\mathrm{IV}}_h - \mathrm{IV}}{\sqrt{2h\widehat{\mathrm{IQ}}_h}} = \frac{\sum_{i=1}^n (\Delta_i X)^2 I_{\{(\Delta_i X)^2 \le r(h)\}} - \int_0^T \sigma_t^2\,\mathrm{d}t}{\sqrt{2h\widehat{\mathrm{IQ}}_h}}$$

can be written as

$$\frac{\sum_{i=1}^n (\Delta_i X_0)^2 - \int_0^T \sigma_t^2\,\mathrm{d}t}{\sqrt{2h\int_0^T \sigma_u^4\,\mathrm{d}u}} \frac{\sqrt{\int_0^T \sigma_u^4}}{\sqrt{\widehat{\mathrm{IQ}}_h}} - \frac{\sum_{i=1}^n (\Delta_i X_0)^2 I_{\{\Delta_i N \neq 0\}}}{\sqrt{2h\widehat{\mathrm{IQ}}_h}}.$$

Since $\int_0^T \sigma_u^4\,\mathrm{d}u/\widehat{\mathrm{IQ}}_h \xrightarrow{P} 1$ and, similarly as in (16), the second term tends to zero in probability, we have the desired result.

Proof of theorem 3.

$$\begin{aligned}\sqrt{n}\sum_{i=1}^n \left(\hat{\gamma}^{(i)} - \gamma^{(i)} I_{\{\Delta_i N \ge 1\}}\right) &= \sqrt{n}\sum_{i=1}^n \Delta_i X I_{\{(\Delta_i X)^2 > r(h),\ \Delta_i N = 0\}} \\ &\quad + \sqrt{n}\sum_{i=1}^n \left(\Delta_i X I_{\{(\Delta_i X)^2 > r(h),\ \Delta_i N = 1\}} - \gamma^{(i)} I_{\{\Delta_i N = 1\}}\right) \\ &\quad + \sqrt{n}\sum_{i=1}^n \left(\Delta_i X I_{\{(\Delta_i X)^2 > r(h),\ \Delta_i N \ge 2\}} - \gamma^{(i)} I_{\{\Delta_i N \ge 2\}}\right).\end{aligned}$$

By theorem 1, a.s. for small h, the first term vanishes. The third term tends to zero in probability, since

$$P\{\sqrt{n}\sum_{i=1}^n \left(\Delta_i X I_{\{(\Delta_i X)^2 > r(h)\}} - \gamma^{(i)}\right) I_{\{\Delta_i N \ge 2\}} \neq 0\} \le P(\cup_{i=1}^n \{\Delta_i N \ge 2\}) \le nO(h^2) = O(h).$$

Therefore, we only need to compute the limit in distribution of the second term. By theorem 1, a.s. for small h it coincides with

$$\operatorname*{dlim}_{h\to 0}\sqrt{n}\sum_{i=1}^{n}\left(\Delta_i X-\gamma^{(i)}\right)I_{\{\Delta_i N=1\}}=\operatorname*{dlim}_{h\to 0}\sqrt{n}\sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s\right)I_{\{\Delta_i N=1\}}. \quad (17)$$

However, since for small h

$$P\left\{\sqrt{n}\sum_{i=1}^{n}\left|\int_{t_{i-1}}^{t_i}a_s\,\mathrm{d}s\right|I_{\{\Delta_i N=1\}}>\epsilon\right\}\le P\left\{\sqrt{n}C(\omega)h^{\mu}N_T>\epsilon\right\}\le P\left\{h^{\mu-0.5}\sqrt{T}C(\omega)N_T>\epsilon\right\},$$

which tends to zero as $h\to 0$, then (17) coincides with

$$\operatorname*{dlim}_{h\to 0}\sqrt{n}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s I_{\{\Delta_i N=1\}},$$

for which we compute the characteristic function

$$E\left[\mathrm{e}^{i\theta\sqrt{n}\sum_i\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s I_{\{\Delta_i N=1\}}}\right]=E\left[\mathrm{e}^{i\theta\sqrt{T}\sum_i\frac{\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s}{\sqrt{h}}I_{\{\Delta_i N=1\}}}\right].$$

Conditionally on σ, for $i=1..n$, $\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s/\sqrt{h}$ are independent Gaussian r.v.s with law $\mathcal{N}(0,\int_{t_{i-1}}^{t_i}\sigma_s^2\,\mathrm{d}s/h)$. Since W and N are independent (Ikeda & Watanabe, 1981), our characteristic function equals

$$\begin{aligned}\Pi_{i=1}^{n}E\left[\mathrm{e}^{i\theta\sqrt{T}\frac{\int_{t_{i-1}}^{t_i}\sigma_s\,\mathrm{d}W_s}{\sqrt{h}}}I_{\{\Delta_i N=1\}}+I_{\{\Delta_i N\neq 1\}}\right]&=\Pi_{i=1}^{n}\left(1+\left(\mathrm{e}^{-\frac{1}{2}\theta^2 T\frac{\int_{t_{i-1}}^{t_i}\sigma_s^2\,\mathrm{d}s}{h}}-1\right)\mathrm{e}^{-\lambda h}\lambda h\right)\\&:=\Pi_{i=1}^{n}(1+\theta_{ni}). \quad (18)\end{aligned}$$

However, $\max_i|\theta_{ni}|\to 0$, $\sum_{i=1}^{n}|\theta_{ni}|\le\lambda$ and

$$\sum_{i=1}^{n}\theta_{ni}=\lambda\,\mathrm{e}^{-\lambda h}\sum_{i=1}^{n}\left(\mathrm{e}^{-\frac{1}{2}\theta^2 T\sigma_{\xi_i}^2}-1\right)h\to\lambda\int_0^T\left(\mathrm{e}^{-\frac{\theta^2 T}{2}\sigma_s^2}-1\right)\mathrm{d}s,$$

where, for each i, ξ_i are suitable points belonging to $]t_{i-1},t_i[$. Therefore (Chung, 1974, p. 199) (18) tends to $\exp\{\lambda\int_0^T(\mathrm{e}^{-\frac{\theta^2 T}{2}\sigma_s^2}-1)\,\mathrm{d}s\}$, which coincides with $E[\exp(-\frac{\theta^2 T}{2}\int_0^T\sigma_s^2\,\mathrm{d}N_s)]$ (Cont & Tankov, 2004, p. 78), the characteristic function of a mixed Gaussian r.v. ηZ where $Z(P)=\mathcal{N}(0,1)$ and $\eta^2=T\int_0^T\sigma_s^2\,\mathrm{d}N_s$.

Lemma 1

Let us consider the sequence $\widehat{\mathrm{IV}}_h, h=\frac{T}{n}, n\in\mathbb{N}$. Under the assumptions of theorem 4 we can find a subsequence h_k for which a.s., for any $\delta>0$, for large k, for all $i=1..n_k$, $n_k=T/h_k$, on $\{(\Delta_i\tilde{J}_2)^2\le 4r(h_k)\}$ we have

$$(\Delta\tilde{J}_{2,s})^2\le\delta+4r(h_k),\quad\forall s\in]t_{i-1},t_i].$$

Proof. For any semimartingale Z (Métivier, 1982, theorem 25.1) we have not only that

$$\sum_{i=1}^{n}(\Delta_i Z)^2\to_P[Z]_T,$$

but, defining $\Pi_n(t)$ as the partition of $[0,t]$ induced by $\{0,t_1,\dots,t_n=T\}$ and defining $R^n(t,Z):=\sum_{t_i\in\Pi_n(t)}(\Delta_i Z)^2$, we also have that there exists a subsequence n_k for which

$$\text{a.s. } R^{n_k}(\cdot,Z)\to[Z]_\cdot \text{ as } k\to\infty, \text{ uniformly in t} \in[0,T]. \tag{19}$$

For $Z\equiv\tilde{J}_2$ we have that the quadratic variation $[Z]_t$ is precisely the sum of the squared sizes of the jumps occurred until t (Cont & Tankov, 2004, p. 266): $[Z]_t=\sum_{s\le t}(\Delta\tilde{J}_{2,s})^2$. This allows us to say that, defining $f_k(t):=R^{n_k}(t,\tilde{J}_2)-\sum_{s\le t}(\Delta\tilde{J}_{2,s})^2$, we have a.s.

$$\sup_{i=1..n}\left|(\Delta_i\tilde{J}_2)^2-\sum_{s\in]t_{i-1},t_i]}(\Delta\tilde{J}_{2,s})^2\right|=\sup_{i=1..n}|f_k(t_i)-f_k(t_{i-1})|\le 2\sup_{t\in[0,T]}|f_k(t)|\to 0,$$

as $k\to\infty$. Therefore, given arbitrary $\delta>0$, for small h_k we have

$$\sup_{i=1..n}\left|(\Delta_i\tilde{J}_2)^2-\sum_{s\in]t_{i-1},t_i]}(\Delta\tilde{J}_{2,s})^2\right|<\delta,$$

and thus

$$\forall i\left|\sum_{s\in]t_{i-1},t_i]}(\Delta\tilde{J}_{2,s})^2\right|-|(\Delta_i\tilde{J}_2)^2|<\left|(\Delta_i\tilde{J}_2)^2-\sum_{s\in]t_{i-1},t_i]}(\Delta\tilde{J}_{2,s})^2\right|<\delta,$$

so that for all i, on $\{(\Delta_i\tilde{J}_2)^2\le 4r(h_k)\}$,

$$\sum_{s\in]t_{i-1},t_i]}(\Delta\tilde{J}_{2,s})^2<\delta+(\Delta_i\tilde{J}_2)^2\le\delta+4r(h_k).$$

In particular $\forall\delta>0$, for sufficiently large k, for all $i=1..n_k$, on $\{(\Delta_i\tilde{J}_2)^2\le 4r(h_k)\}$, we have

$$(\Delta\tilde{J}_{2,s})^2\le\delta+4r(h_k) \text{ for each } s\in]t_{i-1},t_i].$$

Corollary 3

Under the assumptions of theorem 4, as $n\to\infty$,

$$Z_T^{(n)}:=\sum_{i=1}^n(\Delta_i\tilde{J}_2)^2 I_{\{(\Delta_i\tilde{J}_2)^2\le 4r(h)\}}\xrightarrow{P}0.$$

Proof. As a consequence of the previous lemma we find a subsequence n_k such that a.s., for any $\delta>0$, for large k, for all $i=1..n_k$, on $\{(\Delta_i\tilde{J}_2)^2\le 4r(h_k)\}$ there are no jumps bigger in absolute value than $\sqrt{\delta+4r(h_k)}$, that is $\int_{t_{i-1}}^{t_i}\int_{\sqrt{\delta+4r(h_k)}<|x|\le 1}x\mu(\mathrm{d}s,\mathrm{d}x)=0$. Defining the process $Y^{(\delta,h_k)}$ by

$$Y_t^{(\delta,h_k)}:=\int_0^t\int_{|x|\le\sqrt{\delta+4r(h_k)}}x[\mu(\mathrm{d}s,\mathrm{d}x)-\nu(\mathrm{d}x)\,\mathrm{d}s]-t\int_{\sqrt{\delta+4r(h_k)}\le|x|\le 1}x\nu(\mathrm{d}x),$$

we have that $\Delta_i\tilde{J}_2 I_{\{(\Delta_i\tilde{J}_2)^2\le 4r(h_k)\}}$ are in fact those increments of $Y^{(\delta,h_k)}$ which are in absolute value below $2\sqrt{r(h_k)}$, therefore

$$\begin{aligned}\operatorname*{Plim}_{k\to\infty}Z_T^{(n_k)}&=\operatorname*{Plim}_{k\to\infty}\sum_{i=1}^n(\Delta_i\tilde{J}_2)^2 I_{\{(\Delta_i\tilde{J}_2)^2\le 4r(h_k)\}}\le\operatorname*{Plim}_{k\to\infty}\sum_{i=1}^n(\Delta_i Y^{(\delta,h_k)})^2=\operatorname*{Plim}_{k\to\infty}[Y^{(\delta,h_k)}]_T\\&=\operatorname*{Plim}_{k\to\infty}\int_0^T\int_{|x|\le\sqrt{\delta+4r(h_k)}}x^2\mu(\mathrm{d}s,\mathrm{d}x)=\int_0^T\int_{|x|\le\sqrt{\delta}}x^2\mu(\mathrm{d}s,\mathrm{d}x).\end{aligned}$$

Since δ is arbitrary, we can conclude that $\operatorname*{Plim}_{k\to\infty}Z_T^{(n_k)}=0$, as $E[\int_0^T\int_{|x|<\sqrt{\delta}}x^2\mu(\mathrm{d}s,\mathrm{d}x)]=T\sigma^2(\sqrt{\delta})\to 0$, as $\delta\to 0$.

Therefore, we have a subsequence $Z_T^{(n_k)}$ tending to zero in probability as $k \to \infty$, so we can extract a subsequence n_{k_ℓ} converging to zero a.s. However, repeating our reasoning, from each subsequence $Z_T^{(n_m)}$ we can extract a subsequence $Z_T^{(n_{m_v})}$ tending to zero a.s. In conclusion $Z_T^{(n)} \xrightarrow{P} 0$.

Proof of theorem 4. To prove theorem 4 we decompose X into the sum of the FA jump diffusion process X_1 and the infinite activity compensated process $\tilde{J}_2$ of the small jumps. We use corollary 2 for the first term, and we show that the contribution of each $\Delta_i \tilde{J}_2$ to $\widehat{\mathrm{IV}}_h$ is negligible.

Since $X = X_1 + \tilde{J}_2$, we can write

$$\begin{aligned}\left|\sum_{i=1}^n (\Delta_i X)^2 I_{\{(\Delta_i X)^2 \le r(h)\}} - \int_0^t \sigma^2\,\mathrm{d}t\right| &\le \left|\sum_{i=1}^n (\Delta_i X_1)^2 I_{\{(\Delta_i X_1)^2 \le 4r(h)\}} - \int_0^t \sigma^2\,\mathrm{d}t\right| \\ &\quad + \left|\sum_{i=1}^n (\Delta_i X_1)^2 (I_{\{(\Delta_i X)^2 \le r(h)\}} - I_{\{(\Delta_i X_1)^2 \le 4r(h)\}})\right| \\ &\quad + 2\left|\sum_{i=1}^n \Delta_i X_1 \Delta_i \tilde{J}_2 I_{\{(\Delta_i X)^2 \le r(h)\}}\right| \\ &\quad + \left|\sum_{i=1}^n (\Delta_i \tilde{J}_2)^2 I_{\{(\Delta_i X)^2 \le r(h)\}}\right|. \end{aligned} \tag{20}$$

By corollary 2 the first term of the right-hand side tends to zero in probability. We now show that the Plim of each one of the other three terms of the right-hand side is zero as $h \to 0$.

Let us deal with the *second term*:

$$\begin{aligned}&\left|\sum_{i=1}^n (\Delta_i X_1)^2 (I_{\{(\Delta_i X)^2 \le r(h)\}} - I_{\{(\Delta_i X_1)^2 \le 4r(h)\}})\right| \\ &= \left|\sum_{i=1}^n (\Delta_i X_1)^2 (I_{\{(\Delta_i X)^2 \le r(h),\, (\Delta_i X_1)^2 > 4r(h)\}} - I_{\{(\Delta_i X)^2 > r(h),\, \Delta_i X_1)^2 \le 4r(h)\}})\right|. \end{aligned} \tag{21}$$

If $I_{\{(\Delta_i X)^2 \le r(h),\, (\Delta_i X_1)^2 > 4r(h)\}} = 1$, since $2\sqrt{r(h)} - |\Delta_i \tilde{J}_2| < |\Delta_i X_1| - |\Delta_i \tilde{J}_2| \le |\Delta_i X_1 + \Delta_i \tilde{J}_2| \le \sqrt{r(h)}$, then $|\Delta_i \tilde{J}_2| > \sqrt{r(h)}$. Thus a.s.

$$\begin{aligned}\sum_{i=1}^n (\Delta_i X_1)^2 I_{\{(\Delta_i X)^2 \le r(h),\, (\Delta_i X_1)^2 > 4r(h)\}} &\le \sum_{i=1}^n (\Delta_i X_1)^2 I_{\{(\Delta_i \tilde{J}_2)^2 > r(h)\}} \\ &\le 2\sum_{i=1}^n (\Delta_i X_0)^2 I_{\{(\Delta_i \tilde{J}_2)^2 > r(h)\}} + 2\sum_{i=1}^n \left(\sum_{j=1}^{\Delta_i N} \gamma_j\right)^2 I_{\{(\Delta_i \tilde{J}_2)^2 > r(h)\}}. \end{aligned} \tag{22}$$

The first term is a.s. dominated by

$$2\Lambda^2 h \ln\frac{1}{h} \sum_{i=1}^n I_{\{(\Delta_i \tilde{J}_2)^2 > r(h)\}} \xrightarrow{P} 0, \tag{23}$$

as $h \to 0$, since

$$h \ln\frac{1}{h} nP\{(\Delta_i \tilde{J}_2)^2 > r(h)\} \le h \ln\frac{1}{h} n \frac{E[(\Delta_i \tilde{J}_2)^2]}{r(h)} = nh\sigma^2(1) \frac{h \ln\frac{1}{h}}{r(h)} \to 0.$$

Moreover,

$$P\left\{\sum_{i=1}^{n}\left(\sum_{j=1}^{\Delta_i N}\gamma_j\right)^2 I_{\{(\Delta_i\tilde{J}_2)^2>r(h)\}}\neq 0\right\}\leq P\left(\cup_{i=1}^{n}\{\Delta_i N\neq 0, (\Delta_i\tilde{J}_2)^2>r(h)\}\right)$$
$$\leq nP\{\Delta_1 N\neq 0\}\frac{E[(\Delta_1\tilde{J})^2]}{r(h)}=nO(h)\frac{h\sigma^2(1)}{r(h)}\to 0. \tag{24}$$

For the second term of (21) we note that, by theorem 1, for small h on $\{(\Delta_i X_1)\leq 4r(h)\}$ we have, for all $i=1..n$, $\Delta_i N=0$. Therefore, for small h

$$\{(\Delta_i X)^2>r(h), (\Delta_i X_1)^2\leq 4r(h)\}\subset\left\{(\Delta_i X_0+\Delta_i\tilde{J}_2)^2>r(h)\right\}\subset\left\{(\Delta_i X_0)^2>\frac{r(h)}{4}\right\}$$
$$\cup\left\{(\Delta_i\tilde{J}_2)^2>\frac{r(h)}{4}\right\}.$$

However, by (14), a.s., for small h, $I_{\{(\Delta_i X_0)^2>\frac{r(h)}{4}\}}=0\ \forall i=1..n$, and thus

$$\sum_{i=1}^{n}(\Delta_i X_1)^2 I_{\{(\Delta_i X_1)^2>r(h),\,(\Delta_i X_1)^2\leq 4r(h)\}}\leq\sum_{i=1}^{n}(\Delta_i X_0)^2 I_{\{(\Delta_i\tilde{J}_2)^2>\frac{r(h)}{4}\}},$$

which tends to zero as before in (23). Therefore (21) tends to zero in probability as $h\to 0$.

Let us now deal with (half) the Plim of the *third term* on the right-hand side of (20), which coincides with

$$\operatorname*{Plim}_{h\to 0}\sum_{i=1}^{n}\Delta_i X_1\Delta_i\tilde{J}_2 I_{\{|\Delta_i X|\leq\sqrt{r(h)},\,|\Delta_i\tilde{J}_2|\leq 2\sqrt{r(h)}\}}. \tag{25}$$

In fact if

$$|\Delta_i X|\leq\sqrt{r(h)}\text{ and }|\Delta_i\tilde{J}_2|>2\sqrt{r(h)},$$

then

$$2\sqrt{r(h)}-|\Delta_i X_1|<|\Delta_i\tilde{J}_2|-|\Delta_i X_1|\leq|\Delta_i X|\leq\sqrt{r(h)},$$

i.e. $|\Delta_i X_1|>\sqrt{r(h)}$, so that

$$|\Delta_i J_1|+|\Delta_i X_0|>|\Delta_i J_1+\Delta_i X_0|>\sqrt{r(h)},$$

and then

$$\text{either}\quad|\Delta_i J_1|>\frac{\sqrt{r(h)}}{2}\quad\text{or}\quad|\Delta_i X_0|>\frac{\sqrt{r(h)}}{2}. \tag{26}$$

Since a.s., for small h, $I_{\{|\Delta_i X_0|>\frac{\sqrt{r(h)}}{2}\}}=0$, for all $i=1..n$, then

$$P\left\{\sum_{i=1}^{n}|\Delta_i X_1\Delta_i\tilde{J}_2|I_{\{|\Delta_i X|\leq\sqrt{r(h)},\,|\Delta_i\tilde{J}_2|>2\sqrt{r(h)}\}}\neq 0\right\}\leq P\left(\cup_{i=1}^{n}\{|\Delta_i\tilde{J}_2|>2\sqrt{r(h)},\ \Delta_i N\neq 0\}\right), \tag{27}$$

which tends to zero as in (24).

In order to deal now with (25), let us take ω outside the negligible set $\{\underline{\gamma}=0\}$ and note that if $|\Delta_i X|\leq\sqrt{r(h)}$ and $|\Delta_i\tilde{J}_2|\leq 2\sqrt{r(h)}$, then

$$|\Delta_i J_1|-|\Delta_i X_0+\Delta_i\tilde{J}_2|<|\Delta_i X|\leq\sqrt{r(h)},$$

and thus, for small h,

$$|\Delta_i J_1|<\Lambda\sqrt{2h\ln\frac{1}{h}}+|\Delta_i\tilde{J}_2|+\sqrt{r(h)}=O(\sqrt{r(h)}),\quad\forall i=1..n.$$

However, for small h the increment $\Delta_i N$ belongs to $\{0, 1\}$, thus either we have $\Delta_i J_1 = 0$ or $\Delta_i J_1 = \gamma^{(i)} \geq \underline{\gamma}$, and so $\underline{\gamma} < O(\sqrt{r(h)})$. In this second case, we would have a contradiction when h is sufficiently small. Therefore a.s. if $I_{\{|\Delta_i X| \leq \sqrt{r(h)},\, |\Delta_i \tilde{J}_2| \leq 2\sqrt{r(h)}\}} = 1$, then for sufficiently small h we have, for all $i = 1..n$, $\Delta_i N = 0$, and thus (25) is dominated by

$$\begin{aligned}
&\operatorname*{Plim}_{h\to 0} \sum_{i=1}^{n} |\Delta_i X_0 \Delta_i \tilde{J}_2| I_{\{|\Delta_i \tilde{J}_2| \leq 2\sqrt{r(h)}\}} \\
&\quad \leq \operatorname*{Plim}_{h\to 0} \sqrt{\sum_{i=1}^{n} (\Delta_i X_0)^2} \sqrt{\sum_{i=1}^{n} (\Delta_i \tilde{J}_2)^2 I_{\{|\Delta_i \tilde{J}_2| \leq 2\sqrt{r(h)}\}}} = \sqrt{\mathrm{IV}} \operatorname*{Plim}_{h\to 0} \sqrt{Z_T^{(n)}} = 0,
\end{aligned}$$

by the Schwartz inequality and corollary 3.

Finally let us show that the *last term* of the right-hand side of (20) tends to zero in probability. Analogous to (27), the Plim of the last term of (20) coincides with

$$\operatorname*{Plim}_{h\to 0} \sum_{i=1}^{n} (\Delta_i \tilde{J}_2)^2 I_{\{|\Delta_i X| \leq \sqrt{r(h)},\, |\Delta_i \tilde{J}_2| \leq 2\sqrt{r(h)}\}} \leq \operatorname*{Plim}_{h\to 0} \sum_{i=1}^{n} (\Delta_i \tilde{J}_2)^2 I_{\{(\Delta_i \tilde{J}_2)^2 \leq 4r(h)\}} = 0$$

by corollary 3.

Remark for the case of not equally spaced observations. Theorem 4 is still valid if we have not equally spaced observations, as soon as we set $h := \max_i \Delta t_i$. In fact the term $E[I_{\{(\Delta_i \tilde{J}_2)^2 > r(h)\}}]$, which we often use from (23) on, is still negligible, since as $h = \max \Delta t_i \to 0$, $P\{(\Delta_i \tilde{J}_2)^2 > r(h)\} \leq E[(\Delta_i \tilde{J}_2)^2]/r(h) = \Delta t_i \sigma^2(1)/r(h) \leq c\, h/r(h)$. Moreover, repeating the reasoning of lemma 1, we still find n_k such that $\forall \delta > 0$, for sufficiently large k, for all $i = 1..n_k$, on $\{(\Delta_i \tilde{J}_2)^2 \leq 4r(h_k)\}$ we have $(\Delta \tilde{J}_{2,s})^2 \leq \delta + 4r(h_k)$ $\forall s \in]t_{i-1}, t_i]$, so that corollary 3 still holds.