Bernoulli **17**(2), 2011, 781–813
DOI: 10.3150/10-BEJ293

Nonparametric tests for pathwise properties of semimartingales

RAMA CONT[1] and CECILIA MANCINI[2]

[1]*IEOR Department, Columbia University, New York and Laboratoire de Probabilités et Modèles Aléatoires, CNRS-Université Paris VI, France. E-mail: Rama.Cont@columbia.edu*
[2]*Dipartimento di Matematica per le Decisioni, Università di Firenze, Italy.*
E-mail: cecilia.mancini@dmd.unifi.it

We propose two nonparametric tests for investigating the pathwise properties of a signal modeled as the sum of a Lévy process and a Brownian semimartingale. Using a nonparametric threshold estimator for the continuous component of the quadratic variation, we design a test for the presence of a continuous martingale component in the process and a test for establishing whether the jumps have finite or infinite variation, based on observations on a discrete-time grid. We evaluate the performance of our tests using simulations of various stochastic models and use the tests to investigate the fine structure of the DM/USD exchange rate fluctuations and SPX futures prices. In both cases, our tests reveal the presence of a non-zero Brownian component and a finite variation jump component.

Keywords: high frequency data; jump processes; nonparametric tests; quadratic variation; realized volatility; semimartingale

1. Introduction

Continuous-time stochastic models based on *discontinuous semimartingales* have been increasingly used in many applications, such as financial econometrics, option pricing and stochastic control. Some of these models are constructed by adding i.i.d. jumps to a continuous process driven by Brownian motion [16,22], while others are based on purely discontinuous processes which move only through jumps [8,18]. Even within the class of purely discontinuous models, one finds a variety of models with different path properties – finite/infinite jump intensity, finite/infinite variation – which turn out to have an importance in applications, such as optimal stopping [5] and the asymptotic behavior of option prices [9,10]. It is therefore of interest to investigate which class of models – diffusion, jump-diffusion or pure-jump – is the most appropriate for a given data set. Nonparametric procedures have been recently proposed for investigating the presence of jumps [2,6,17] and studying some fine properties of the jumps [3,4,25,26] in a signal. Here, we address related, but different, issues: for a semimartingale whose jump component is a Lévy process, we propose a test for the presence of a continuous martingale component in the price process, which allows us to discriminate between pure-jump and jump-diffusion models, and a test for determining whether the jump component has finite or infinite variation. Our tests are based on a nonparametric threshold estimator [20] for the integrated variance (defined as the continuous component of the quadratic variation) based on observations on a discrete-time

1350-7265

grid. Without imposing restrictive assumptions on the continuous martingale component, we obtain a central limit theorem for this threshold estimator (Section 3) and use it to design our tests (Section 4).

Using simulations of stochastic models commonly used in finance, we check the performance of our tests for realistic sample sizes (Section 5). Applied to time series of the DM/USD exchange rate and SPX futures prices (Section 6), our tests reveal, in both cases, the presence of a non-zero Brownian component, combined with a finite variation jump component. These results suggest that these asset prices may be modeled as the sum of a Brownian martingale and a jump component of finite variation.

2. Definitions and notation

We consider a semimartingale $(X_t)_{t\in[0,T]}$, defined on a (filtered) probability space $(\Omega, (\mathcal{F}_t)_{t\in[0,T]}, \mathcal{F}, P)$ with paths in $D([0,T],\mathbb{R})$, driven by a (standard) Brownian motion W and a pure-jump Lévy process L:

$$X_t = x_0 + \int_0^t a_s\,\mathrm{d}s + \int_0^t \sigma_s\,\mathrm{d}W_s + L_t, \qquad t \in]0,T], \tag{1}$$

where a, σ are adapted processes with right-continuous paths with left limits (cadlag processes), such that (1) admits a unique strong solution X on $[0,T]$ which is adapted and cadlag [11]. L has Lévy measure ν and may be decomposed as $L_t = J_t + M_t$, where

$$J_t := \int_0^t\int_{|x|>1} x\mu(\mathrm{d}x,\mathrm{d}s) = \sum_{\ell=1}^{N_t}\gamma_\ell, \qquad M_t := \int_0^t\int_{|x|\le 1} x[\mu(\mathrm{d}x,\mathrm{d}s) - \nu(\mathrm{d}x)\,\mathrm{d}t]. \tag{2}$$

J is a compound Poisson process representing the "large" jumps of X, μ is a Poisson random measure on $[0,T]\times\mathbb{R}$ with intensity measure $\nu(\mathrm{d}x)\,\mathrm{d}t$, N is a Poisson process with intensity $\nu(\{x, |x|>1\}) < \infty$, γ_ℓ are i.i.d. and independent of N and the martingale M is the compensated sum of small jumps of L. We will define $\mu(\mathrm{d}x,\mathrm{d}t) - \nu(\mathrm{d}x)\,\mathrm{d}t =: \tilde{\mu}(\mathrm{d}x,\mathrm{d}t)$, the compensated Poisson random measure associated to μ. We allow for the *infinite activity* (IA) case $\nu(\mathbb{R}) = \infty$, where small jumps of L occur infinitely often. For a semimartingale Z, we denote by $\Delta_i Z = Z_{t_i} - Z_{t_{i-1}}$ its increments and by $\Delta Z_t = Z_t - Z_{t-}$ its jump at time t. The *Blumenthal–Getoor (BG) index* of L, defined as

$$\alpha := \inf\left\{\delta \ge 0, \int_{|x|\le 1} |x|^\delta \nu(\mathrm{d}x) < +\infty\right\} \le 2,$$

measures the degree of *activity* of small jumps. A compound Poisson process has $\alpha = 0$, while an α-stable process has BG index equal to $\alpha \in]0,2[$. The gamma process and the variance gamma (VG) process are examples of infinite activity Lévy processes with $\alpha = 0$. A pure-jump Lévy process with BG index $\alpha < 1$ has paths with *finite variation*, while for $\alpha > 1$, the sample paths have *infinite variation* a.s. When $\alpha = 1$, the paths may have either finite or infinite variation [7]. The normal inverse Gaussian process (NIG) and the generalized hyperbolic Lévy motion (GHL) have infinite variation and $\alpha = 1$. Tempered stable processes [8,10] allow for $\alpha \in [0, 2[$. We call $IV = \int_0^T \sigma_u^2 \, \mathrm{d}u$ the *integrated variance* of X and $IQ = \int_0^T \sigma_u^4 \, \mathrm{d}u$ the *integrated quarticity* of X, and we write

$$X_{0t} = \int_0^t a_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s, \qquad X_{1t} = X_{0t} + J_t.$$

We will use the following assumption.

Assumption A1.

$$\exists \alpha \in [0, 2] \qquad \int_{|x| \le \varepsilon} x^2 \nu(\mathrm{d}x) \sim \varepsilon^{2-\alpha} \qquad \textit{as } \varepsilon \to 0, \tag{3}$$

where $f(h) \sim g(h)$ *means that* $f(h) = \mathrm{O}(g(h))$ *and* $g(h) = \mathrm{O}(f(h))$ *as* $h \to 0$.

This assumption implies that α is the BG index of L. A1 is satisfied if, for instance, ν has a density which behaves as $\frac{K_\pm}{|x|^{1+\alpha}}$ when $x \to 0\pm$, where $K_\pm > 0$. In particular, A1 holds for all Lévy processes commonly used in finance [10]: NIG, variance gamma, tempered stable processes or generalized hyperbolic processes.

Typically, we observe X_t in the form of a discrete record $\{x_0, X_{t_1}, \ldots, X_{t_{n-1}}, X_{t_n}\}$ on a time grid $t_i = ih$ with $h = T/n$. Our goal is to provide, given such a discrete observations, nonparametric tests for:

- detecting the presence of a continuous martingale component in the price process;
- analyzing the qualitative nature of the jump component, that is, whether it has finite or infinite variation.

3. Central limit theorem for a threshold estimator of integrated variance

The "realized variance" $\sum_{i=1}^n (\Delta_i X)^2$ of the semimartingale X converges in probability [24] to

$$[X]_T := \int_0^T \sigma_t^2 \, \mathrm{d}t + \int_0^T \int_{\mathbb{R}-\{0\}} x^2 \mu(\mathrm{d}x, \mathrm{d}s).$$

A *threshold estimator* [19,20] of the integrated variance $IV = \int_0^T \sigma_t^2 \, \mathrm{d}t$ is based on the idea of summing only some of the squared increments of X, those whose absolute value is smaller than some *threshold* r_h:

$$\hat{IV}_h := \sum_{i=1}^n (\Delta_i X)^2 I_{\{(\Delta_i X)^2 \le r_h\}}. \tag{4}$$

The term $\int_0^T \int_{\mathbb{R}-\{0\}} x^2 \mu(\mathrm{d}x, \mathrm{d}s)$, due to jumps, vanishes as $h \to 0$ for an appropriate choice of the threshold. P. Lévy's law for the modulus of continuity of the Brownian paths implies that

$$P\left(\lim_{h\to 0} \sup_{i\in\{1,\dots,n\}} \frac{|\Delta_i W|}{\sqrt{2h \ln 1/h}} \le 1\right) = 1$$

and allows such a threshold to be chosen. It is shown in [20], Corollary 2, Theorem 4, that, under the above assumptions, if we choose a deterministic threshold r_h such that

$$\lim_{h\to 0} r_h = 0 \quad \text{and} \quad \lim_{h\to 0} \frac{h \ln h}{r_h} = 0, \tag{5}$$

then $\hat{IV}_h \xrightarrow{P} IV$ as $h \to 0$. If the jumps have finite intensity, then the thresholding procedure allows as $h \to 0$, a jump to be detected in $]t_{i-1}, t_i]$. In fact, since a and σ are cadlag (or caglad), their paths are a.s. bounded on $[0, T]$, so

$$\begin{aligned} &\limsup_{h\to 0} \frac{\sup_i |\int_{t_{i-1}}^{t_i} a_s(\omega)\,\mathrm{d}s|}{h} \le A(\omega) < \infty \quad \text{and} \\ &\limsup_{h\to 0} \frac{\sup_i |\int_{t_{i-1}}^{t_i} \sigma_s^2(\omega)\,\mathrm{d}s|}{h} \le \Sigma(\omega) < \infty \qquad \text{a.s.} \end{aligned} \tag{6}$$

It follows from [20] that

$$\text{a.s.} \qquad \sup_i \frac{|\int_{t_{i-1}}^{t_i} a_s\,\mathrm{d}s + \int_{t_{i-1}}^{t_i} \sigma_s\,\mathrm{d}W_s|}{\sqrt{2h \log 1/h}} \le A\sqrt{h} + \sqrt{\Sigma} + 1 := \Lambda. \tag{7}$$

Since realistic values of σ for asset prices belong to $[0.1, 0.8]$ (in annual units), we have that for small h, the r.v. Λ has order of magnitude of 1, thus, in the finite jump intensity case, a.s. for sufficiently small h, $(\Delta_i X)^2 > r_h > 2h \log \frac{1}{h}$ indicates the presence of jumps in $]t_{i-1}, t_i]$.

When L has infinite activity, $\sum_{i=1}^n (\Delta_i X)^2 I_{\{(\Delta_i X)^2 \le r_h\}}$ behaves like $\sum_{i=1}^n (\Delta_i X)^2 \times I_{\{\Delta_i N = 0, |\Delta_i M| \le 2\sqrt{r_h}\}}$ for small h (Lemma A.2). Moreover, for any $\delta > 0$, the jumps contributing to the increments $\Delta_i X$ such that $(\Delta_i X)^2 \le r_h$ for small h have size smaller than $c\sqrt{r_h + \delta}$ ([20], Lemma 1), so their contribution vanishes when $h \to 0$. Note that $r_h = ch^\beta$ satisfies condition (5) for any $\beta \in]0, 1[$ and any constant c. Since $\sqrt{2}\sigma \simeq 1$ in most applications, we use $c = 1$. Define

$$\eta^2(\varepsilon) := \int_{|x|\le\varepsilon} x^2 \nu(\mathrm{d}x), \qquad d(\varepsilon) := \int_{\varepsilon<|x|\le 1} x \nu(\mathrm{d}x). \tag{8}$$

Let us remark that if $\lim_{h\to 0} r_h = 0$, then, by A1, we have, as $h \to 0$,

$$\eta^2\big(2\sqrt{r_h}\big) = \int_{|x|\le 2\sqrt{r_h}} x^2 \nu(\mathrm{d}x) \sim r_h^{1-\alpha/2}, \qquad \int_{|x|\le 2\sqrt{r_h}} x^k \nu(\mathrm{d}x) \sim r_h^{(k-\alpha)/2},$$
$$k = 3, 4, \tag{9}$$

$$\int_{2\sqrt{r_h}<|x|\le 1} x\nu(\mathrm{d}x) \sim \bigl[c + r_h^{(1-\alpha)/2}\bigr] I_{\{\alpha\neq 1\}} + \left[\ln \frac{1}{2\sqrt{r_h}}\right] I_{\{\alpha=1\}},$$

$$\int_{2\sqrt{r_h}<|x|\le 1} \nu(\mathrm{d}x) \sim r_h^{-\alpha/2},$$

where α is the BG index of L. The following lemma, proved in the Appendix, states that under (5), each increment $\Delta_i M$ such that $|\Delta_i M| \le 2\sqrt{r_h}$ only contains jumps of magnitude less than $2\sqrt{r_h}$ if $\alpha \le 1$, or smaller than $2h^{1/(2\alpha)} \log^{1/(2\alpha)} \frac{1}{h}$ if $\alpha > 1$.

Lemma 3.1. *Define, for $h > 0$, $v_h := h^{1/(2\alpha)} \log^{1/(2\alpha)} \frac{1}{h}$. Under (5). there exists a sequence $h_k = T/n_k$ tending to zero as $k \to \infty$ such that, for k_0 sufficiently large and $h \in \{h_k, k \ge k_0\}$:*

(i) *if $\alpha \le 1$, then for all $i = 1, \dots, n$,*

$$\begin{aligned}&\Delta_i M I_{\{(\Delta_i M)^2 \le 4r_h\}}\\ &\quad= \left(\int_{t_{i-1}}^{t_i} \int_{|x|\le 2\sqrt{r_h}} x\tilde{\mu}(\mathrm{d}x, \mathrm{d}t) - \int_{t_{i-1}}^{t_i} \int_{2\sqrt{r_h}<|x|\le 1} x\nu(\mathrm{d}x)\,\mathrm{d}t\right) I_{\{(\Delta_i M)^2 \le 4r_h\}} \qquad a.s.;\end{aligned}$$

(ii) *if $\alpha > 1$, then for all $i = 1, \dots, n$, we have*

$$\begin{aligned}&\Delta_i M I_{\{(\Delta_i M)^2 \le 4r_h\}}\\ &\quad= \left(\int_{t_{i-1}}^{t_i} \int_{|x|\le 2v_h} x\tilde{\mu}(\mathrm{d}x, \mathrm{d}t) - \int_{t_{i-1}}^{t_i} \int_{2v_h<|x|\le 1} x\nu(\mathrm{d}x)\,\mathrm{d}t\right) I_{\{(\Delta_i M)^2 \le 4r_h\}} \qquad a.s.\end{aligned}$$

Remark 3.2. Note that $v_h \le r_h^{1/4}$ so that in the case (ii) above ($\alpha > 1$), for all $i = 1, \dots, n$, the jumps of M on $\{(\Delta_i M)^2 \le 4r_h\}$ are bounded by $r_h^{1/4}$.

Definition. *Define*

$$\begin{aligned} L_t^{(h)} &:= \int_0^t \int_{|x|\le 2\sqrt[4]{r_h}} x\tilde{\mu}(\mathrm{d}x, \mathrm{d}t) - \int_0^t \int_{2\sqrt[4]{r_h}<|x|\le 1} x\nu(\mathrm{d}x)\,\mathrm{d}t,\\ \Delta_i M^{(h)} &:= \int_{t_{i-1}}^{t_i} \int_{|x|\le 2\sqrt[4]{r_h}} x\tilde{\mu}(\mathrm{d}x, \mathrm{d}t). \end{aligned} \tag{10}$$

By Lemma 3.1, on a subsequence, a.s. for sufficiently small h, $\forall i = 1, \dots, n$, on $\{(\Delta_i M)^2 \le 4r_h\}$, we have

$$\Delta_i M = \Delta_i L^{(h)} = \Delta_i M^{(h)} - hd\bigl(2\sqrt[4]{r_h}\bigr). \tag{11}$$

$\Delta_i M^{(h)}$ is the compensated sum of jumps smaller in absolute value than $2\sqrt[4]{r_h}$, while $hd(2\sqrt[4]{r_h})$ is the compensator of the (missing) jumps larger than $2\sqrt[4]{r_h}$.

In [20], a central limit theorem for $\hat{IV}_h$ was shown in the case of finite intensity jumps and cadlag adapted σ. Theorem 3.5 extends this to the case of infinite activity without extra assumptions on σ. In particular, when $\alpha < 1$, the error $\hat{IV}_h - IV$ has the same rate of convergence and asymptotic variance as in the case of finite intensity jumps. The following proposition gives the asymptotic variance of $(\hat{IV}_h - IV)/\sqrt{2h}$ when $\alpha < 1$.

Proposition 3.3. *If $r_h = h^\beta$ with $1 > \beta > \frac{1}{2-\alpha/2} \in [1/2, 1[$, then, as $h \to 0$,*

$$\hat{IQ}_h := \frac{\sum_i (\Delta_i X)^4 I_{\{(\Delta_i X)^2 \le r_h\}}}{3h} \xrightarrow{P} IQ = \int_0^T \sigma_t^4 \, \mathrm{d}t.$$

The following result will be used to prove Theorem 3.5.

Theorem 3.4. *Under Assumption A1, as $h \to 0$,*

$$\frac{\sum_{i=1}^n (\int_{t_{i-1}}^{t_i} \int_{|x|\le\varepsilon} x \tilde{\mu}(\mathrm{d}x, \mathrm{d}t) - \int_{t_{i-1}}^{t_i} \int_{|x|\in]\varepsilon,1]} x \nu(\mathrm{d}x)\,\mathrm{d}t)^2 - T\ell_{2,h}\varepsilon^{2-\alpha} - T\ell_{1,h}^2 h \varepsilon^{2-2\alpha} I_{\{\alpha\neq 1\}}}{\sqrt{T}\sqrt{\ell_{4,h}}\varepsilon^{2-\alpha/2}} \tag{12}$$

$$\xrightarrow{d} N(0,1),$$

where $\varepsilon = h^u$, $0 < u \le 1/2$, $\ell_{j,h} = \int_{|x|\le\varepsilon} x^j \nu(\mathrm{d}x)/\varepsilon^{j-\alpha}$ for $j = 2, 4$ and $\ell_{1,h} = \int_{\varepsilon<|x|\le 1} x\nu(\mathrm{d}x)/ [(c + \varepsilon^{1-\alpha}) I_{\{\alpha\neq1\}} + \ln \frac{1}{2\varepsilon} I_{\{\alpha=1\}}]$ tend to non-zero constants depending on ν.

We are now ready to state our central limit theorem for the estimator $\hat{IV}_h$. A sequence (X_n) is said to converge stably in law to a random variable X (defined on an extension $(\Omega', \mathcal{F}', P')$ of the original probability space) if $\lim E[Uf(X_n)] = E'[Uf(X)]$ for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$ and all bounded random variables U. This is obviously stronger than convergence in law [15].

Theorem 3.5. *Assume* A1 *and $\sigma \not\equiv 0$; choose $r_h = h^\beta$ with $\beta > \frac{1}{2-\alpha/2} \in [1/2, 1[$. Then:*

(a) *if $\alpha < 1$, we have, with $\xrightarrow{st}$ denoting stable convergence in law,*

$$\frac{\hat{IV}_h - IV}{\sqrt{2h\hat{IQ}_h}} \xrightarrow{st} N(0,1); \tag{13}$$

(b) *if $\alpha \ge 1$, then*

$$\frac{\hat{IV}_h - IV}{\sqrt{2h\hat{IQ}_h}} \xrightarrow{a.s.} +\infty.$$

Remark. For $\alpha < 1$, Jacod [13], Theorem 2.10(i), has shown a related central limit result for the threshold estimator of IV, where L is a semimartingale, but under the additional assumption that σ is an Itô semimartingale. The proof of Theorem 3.5 in the case $\alpha < 1$ does not rely on [13], Theorem 2.10(i). An alternative proof under the Itô semimartingale assumption for σ could combine the results [20] with [13], Theorem 2.10(i), in that

$$\frac{\hat{IV} - IV}{\sqrt{h}} = \frac{\hat{IV}(X_1) - IV}{\sqrt{h}} + \frac{\hat{IV}(M)}{\sqrt{h}} + \frac{\sum_{i=1}^n (\Delta_i X_1)^2 (I_{\{(\Delta_i X)^2 \le r_h\}} - I_{\{(\Delta_i X_1)^2 \le r_h\}})}{\sqrt{h}} + \frac{\sum_{i=1}^n (\Delta_i M)^2 (I_{\{(\Delta_i X)^2 \le r_h\}} - I_{\{(\Delta_i M)^2 \le r_h\}})}{\sqrt{h}} + 2\frac{\sum_{i=1}^n \Delta_i X_1 \Delta_i M I_{\{(\Delta_i X)^2 \le r_h\}}}{\sqrt{h}},$$

where

$$\hat{IV}(X_1) \doteq \sum_{i=1}^n (\Delta_i X_1)^2 I_{\{(\Delta_i X_1)^2 \le r_h\}}, \qquad \hat{IV}(M) \doteq \sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i M)^2 \le r_h\}}.$$

The first term converges stably in law by [20], the second one converges stably to zero by [13], Theorem 2.10(i). That the remaining terms are negligible requires some further work (see the proof of Theorem 3.5).

4. Statistical tests

4.1. Test for the presence of a continuous martingale component

We now use the above results to design a test to detect the presence of a continuous martingale component $\int_0^t \sigma_t \, \mathrm{d}W_t$, given discretely recorded observations. Our test is feasible in the case where L has BG index $\alpha < 1$, that is, the jumps are of finite variation (see Section 4.2). The test proceeds as follows. First, we choose a coefficient $\beta \in [1/2, 1[$ close to 1. If we have an estimate $\hat{\alpha}$ of the BG index [3,25,26], then we may choose $\beta > \frac{1}{2-\hat{\alpha}}$ (recall that $\frac{1}{2-\alpha} \in [1/2, 1[$). We choose a threshold $r_h = h^\beta$ and use the estimator $\hat{IQ}_h$ of the integrated quarticity defined in Proposition 3.3. We have shown in Theorem 3.5 that, when $\sigma \not\equiv 0$ in the case $\alpha < 1$, the estimator $\hat{IV}_h$ is asymptotically Gaussian as $h \to 0$. However, if $\sigma \equiv 0$, then both the numerator and the denominator of (13) tend to zero. To handle this case, we add an i.i.d. noise term:

$$\Delta_i X^v := \Delta_i X + v\sqrt{h} Z_i, \qquad Z_i \overset{\text{i.i.d.}}{\sim} N(0, 1).$$

As $h \to 0$,

$$\sum_{i=1}^n (\Delta_i X^v)^2 \xrightarrow{P} [X^v]_T = \int_0^T \sigma_s^2 \, \mathrm{d}s + v^2 T + T \int_{\mathbb{R}-\{0\}} x^2 \mu(\mathrm{d}x, \mathrm{d}s)$$

and $I_{\{(\Delta_i X^v)^2 \le r_h\}}$ removes the jumps of X^v so that under the assumptions of Theorem 3.5, as $h \to 0$,

$$\hat{IV}_h^v := \sum_{i=1}^n (\Delta_i X^v)^2 I_{\{(\Delta_i X^v)^2 \le r_h\}} \xrightarrow{P} \int_0^T \sigma_s^2 \, \mathrm{d}s + v^2 T.$$

Under the null hypothesis $\sigma \equiv 0$, we have $\hat{IV}_h^v \xrightarrow{P} v^2 T$, $\hat{IQ}_h^v := \sum_i (\Delta_i X^v)^4 I_{\{(\Delta_i X^v)^2 \le r_h\}}/(3h) \xrightarrow{P} v^4 T$ and

$$U_h := \frac{\hat{IV}_h^v - v^2 T}{\sqrt{2h\hat{IQ}_h^v}} \xrightarrow{st} \mathcal{N}. \tag{14}$$

Note that if, on the contrary, $\sigma \not\equiv 0$, then we have that the limit in probability of $\hat{IV}_h^v$ is strictly larger than $v^2 T$ and, by Lemma A.2, passing to a subsequence, a.s.

$$\begin{aligned}\lim_{h\to 0} h\hat{IQ}_h^v &= \frac{1}{3}\lim_{h\to 0}\sum_i (\Delta_i X^v)^4 I_{\{(\Delta_i X^v)^2\le r_h\}} = \frac{1}{3}\lim_{h\to 0}\sum_i (\Delta_i X^v)^4 I_{\{\Delta_i N=0,(\Delta_i M)^2\le 2r_h\}}\\ &\le \frac{1}{3}\lim_{h\to 0}\sum_i \big(\Delta_i X_0 + \Delta_i M + v\sqrt{h}Z_i\big)^4 I_{\{(\Delta_i M)^2\le 2r_h\}}\\ &\le \frac{c}{3}\lim_{h\to 0}\sum_i (\Delta_i X_0)^4 + \frac{c}{3}\lim_{h\to 0}\sum_i (\Delta_i M)^4 I_{\{(\Delta_i M)^2\le 2r_h\}} + \frac{c}{3}\lim_{h\to 0}\sum_i \big(v\sqrt{h}Z_i\big)^4.\end{aligned}$$

Using the facts that $\lim_{h\to 0}\sum_i (\Delta_i M)^4 I_{\{(\Delta_i M)^2\le 2r_h\}} \le \lim_{h\to 0} 2r_h \sum_i (\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 2r_h\}} = 0$, by (44), $\sum_i (\Delta_i X_0)^4/h \xrightarrow{P} c\int_0^T \sigma_s^4\,\mathrm{d}s$ and $\sum_i (v\sqrt{h}Z_i)^4/h \xrightarrow{a.s.} cv^4$, we have, as $h \to 0$, $h\hat{IQ}_h^v \xrightarrow{P} 0$. Therefore, under the alternative $(H_1)\sigma \not\equiv 0$, $U_h \to +\infty$ and $P\{|U_h| > 1.96\} \to 1$, so the test is consistent.

Local power of the test. To investigate the local power of the test U_h, we consider a sequence of alternatives $(H_1^h)\sigma = \sigma^h$, where $\sigma^h \downarrow 0$. We denote by $\hat{IQ}_{\sigma^h}^v, U_{\sigma^h}$ the statistics analogous to $\hat{IQ}_h^v, U_h$, but constructed from $X_t^h = x_0 + \int_0^t a_s\,\mathrm{d}s + \int_0^t \sigma_s^h\,\mathrm{d}W_s + L_t, t \in]0, T]$. In the case of constant σ and σ^h, and finite jump intensity, using standard results on convergence of sums of a triangular array [14], Lemmas 4.1 and 4.3, we have

$$\hat{IQ}_{\sigma^h}^v \xrightarrow{\text{ucp}} v^4 T, \qquad U_{\sigma^h} \xrightarrow{d} \lim_{h\to 0}\frac{(\sigma^h)^2}{\sqrt{h}}T + \sqrt{2}v^2 Z_T,$$

where $\xrightarrow{\text{ucp}}$ denotes uniform convergence in probability on compacts subsets of $[0, T]$ [24] and Z is a standard Brownian motion. So, either U_{σ^h} tends in distribution to $c + \sqrt{2}v^2 Z_T$, if $\sigma^h = \mathrm{O}(h^{1/4})$, or $U_{\sigma^h} \to \infty$, if $h^{1/4} = \mathrm{o}(\sigma^h)$. Thus, if c is a (possibly zero) constant, we have:

$$\text{if } \frac{\sigma^h}{h^{1/4}} \to c, \qquad \text{then } P\{U_{\sigma^h} > 1.64|H_1^h\} \to P\left\{Z_1 > \frac{1.64 - c^2 T}{\sqrt{2T}v^2}\right\};$$

$$\text{if } \frac{\sigma^h}{h^{1/4}} \to +\infty, \qquad \text{then } P\{U_{\sigma^h} > 1.64|H_1^h\} \to 1.$$

For values of v in Section 5, we have $1.64/\sqrt{2T}v^2 = \mathrm{O}(10^8)$ and thus the local power of the test is small if $\sigma^h = \mathrm{O}(h^{1/4})$.

4.2. Testing whether the jump component has finite variation

To construct a test for discriminating $\alpha < 1$ from $\alpha \geq 1$, Theorem 3.5 suggests the use of $(\hat{IV}_h - IV)/\sqrt{2h\hat{IQ}_h}$, but this requires knowing the process σ to compute IV. We propose a feasible alternative. Consider, instead, the estimator

$$\hat{H}_h := \sum_{i=1}^{n} \Delta_i X I_{\{(\Delta_i X)^2 > r_h\}} = X_T - \sum_{i=1}^{n} \Delta_i X I_{\{(\Delta_i X)^2 \leq r_h\}}.$$

Proposition 4.1. *When $\alpha < 1$, $\hat{H}_h$ is a consistent estimator of $J_T + mT$, $m := \int_{-1}^{1} x\nu(\mathrm{d}x)$.*

Consider $Z_i = \Delta_i W^v$, where W^v is a Wiener process independent of W, L, and define

$$\Delta_i \hat{H}^v := \Delta_i X I_{\{(\Delta_i X)^2 > r_h\}} + v\sqrt{h} Z_i \quad \text{and} \quad H_T^v := J_T + mT + vW_T^v.$$

Under the null hypothesis $\alpha < 1$,

$$\hat{IV}_h^{H^v} := \sum_i (\Delta_i \hat{H}^v)^2 I_{\{(\Delta_i \hat{H}^v)^2 \leq r_h\}}$$

is an estimator of the integrated variance $v^2 T$ of H^v, so, under the null hypothesis (H_0) $\alpha < 1$, we can find $\beta > \frac{1}{2-\alpha} \in]\frac{1}{2}, 1[$ such that

$$U_h^{(\alpha)} := \frac{\hat{IV}_h^{H^v} - v^2 T}{\sqrt{2h\hat{IQ}_h^{H^v}}} \xrightarrow{d} N(0, 1), \tag{15}$$

where $\hat{IQ}_h^{H^v} := \frac{1}{3h} \sum_i (\Delta_i \hat{H}^v)^4 I_{\{(\Delta_i \hat{H}^v)^2 \leq r_h\}}$ and $r_h = h^\beta$. In particular, $P\{|U_h^{(\alpha)}| > 1.96\} \to 5\%$.

If, on the contrary, $\alpha \geq 1$, then reasoning as in Theorem 3.5, for any $\beta \in]0, 1[$, we have $U_h^{(\alpha)} \xrightarrow{P} +\infty$, so the test is consistent. If $|U_h^{(\alpha)}| > 1.96$, then we reject (H_0) $\alpha < 1$ at the 95% confidence level.

Remark. To apply this test, we first need to decide whether $\alpha < 1$, using the previously described test.

5. Numerical experiments

5.1. Testing the finite variation of the jump component

We simulate n increments $\Delta_i X$ of a process $X = \sigma W + L$, where L is a symmetric α-stable Lévy process, $\sigma = 0.2$. We generate 1000 independent samples containing n increments each

Table 1. Testing for finite variation of jumps: α-stable process plus Brownian motion. *pct* is the percentage of outcomes where $|U_{(j)}^{(\alpha)}| > 1.96$

n	h	v	α	pct	α	pct
1000	5 min	0.000001	0.6	0.067	1.6	0.439
1000	5 min	0.0001	0.6	0.056	1.6	0.407
1000	5 min	0.01	0.6	0.047	1.6	0.250
1000	5 min	0.1	0.6	0.053	1.6	0.726
1000	1 min	0.0001	0.6	0.049	1.6	0.241
1000	1 hour	0.0001	0.6	0.051	1.6	0.875
1000	1 day	0.0001	0.6	0.066	1.6	0.984
100	5 min	0.0001	0.6	0.065	1.6	0.137
10 000	5 min	0.0001	0.6	0.065	1.6	0.928

and compute $U_h^{(\alpha)}$ as in (15) for a range of values of v, h (1 minute, 5 minutes, 1 hour, 1 day) and number of observations n. Table 1 reports the percentage (*pct*) of outcomes where $|U_{h(j)}^{(\alpha)}| > 1.96$, $j = 1, \dots, 1000$, for threshold exponent $\beta = 0.999$. Note that with $n = 1000$ and h equal to five minutes ($h = 1/(252 \times 84)$), we have $T < 1$ year; for $\alpha = 0.6$, the lower bound for β is $\frac{1}{2-\alpha} = 0.71$; when $n = 1000$, $h = 1/(84 \times 252)$ and the BG index of L is 0.6 (resp., 1.6), the ratio of $v = 10^{-4}$ to the standard deviation of the increments $\Delta_i X$ is 0.074 (resp., 0.022). The test results are observed to be reliable if we use $n = 10\,000$ observations, a time resolution of five minutes and $v = 10^{-4}$. In fact, when the data-generating process has BG index 0.6, the test leads us to accept the hypothesis (H_0) $\alpha < 1$ in about 94 cases out of 100. On the contrary, when the process has BG index 1.6, the test tells us to reject (H_0) in 92 cases out of 100.

5.2. Test for the presence of a Brownian component

We simulate 1000 independent paths of a process $X_t = \int_0^t \sigma_u \, \mathrm{d}W_u + L$, for different Lévy processes L and constant or stochastic σ, on a time grid with n steps. We take threshold $r_h = h^{0.999}$. For each trial $j = 1, \dots, 1000$, we compute $U_{h(j)}$ given in (14) and report the percentage (*pct*) of cases where $|U_{h(j)}| > 1.96$.

Example 5.1 (Brownian motion plus compound Poisson process, BG index $\alpha = 0$). We consider here constant σ and $L = \sum_{i=1}^{N_t} B_i$, a compound Poisson process with i.i.d. $N(0, 0.6^2)$ sizes of jump and jump intensity $\lambda = 5$ (as in [1]). Table 2 illustrates the performance of our test for various time steps h, numbers of observations n and noise levels v: Note that when $\sigma = 0$ (resp., 0.2), $n = 1000$ and $h = 1/(84 \times 252)$ the ratio of $v = 10^{-4}$ to the standard deviation of the returns $\Delta_i X$ equals 0.007 (resp., 0.052).

We find that the test is reliable for values $n = 1000$, $h = 5$ minutes and $v = 10^{-4}$ since it correctly accepts (H_0) in 95 cases out of 100 and rejects (H_0) in all cases when it is false.

Table 2. Testing for the presence of a Brownian component: case of Brownian motion plus compound Poisson jumps (Example 5.1)

n	h	v	σ	*pct*	σ	*pct*
1000	5 min	0.000001	0	0.043	0.2	1
1000	5 min	0.0001	0	0.048	0.2	1
1000	5 min	0.01	0	0.054	0.2	1
1000	5 min	0.1	0	0.041	0.2	1
1000	1 min	0.0001	0	0.047	0.2	1
1000	1 hour	0.0001	0	0.054	0.2	1
1000	1 day	0.0001	0	0.082	0.2	1
100	5 min	0.0001	0	0.065	0.2	1
10 000	5 min	0.0001	0	0.049	0.2	1

Example 5.2 (Brownian motion plus α-stable jumps: $\alpha \in]0, 2[$). Here, L is a symmetric α-stable Lévy process and σ is constant. The results in Table 3 confirm the satisfactory performance of the test when $\alpha = 0.3 < 1$ for $n = 1000$, $h = 5$ minutes and $v = 10^{-4}$.

Table 4, for the case $\alpha = 1.2 > 1$, confirms that we cannot rely on the test results in this case: even when $\sigma \equiv 0$, the statistic U_h diverges if $\alpha \geq 1$.

The main point here is that we may use a model-free choice of threshold.

Example 5.3 (Stochastic volatility plus variance gamma jumps: $\alpha = 0$). Let us now consider a model X with stochastic volatility σ_t, correlated with the Brownian motion driving X and with jumps given by an independent variance gamma process:

$$\mathrm{d}X_t = (\mu - \sigma_t^2/2)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t^{(1)} + \mathrm{d}L_t,$$

Table 3. Testing for the presence of a Brownian component: case of Brownian motion plus α-stable Lévy process with $\alpha = 0.3$ (Example 5.2)

n	h	v	σ	*pct*	σ	*pct*
1000	5 min	0.000001	0	0.042	0.2	1
1000	5 min	0.0001	0	0.026	0.2	1
1000	5 min	0.01	0	0.054	0.2	1
1000	5 min	0.1	0	0.053	0.2	1
1000	1 min	0.0001	0	0.046	0.2	1
1000	1 hour	0.0001	0	0.140	0.2	1
1000	1 day	0.0001	0	0.805	0.2	1
100	5 min	0.0001	0	0.056	0.2	1
10 000	5 min	0.0001	0	0.165	0.2	1

Table 4. Testing for the presence of a Brownian component: case of Brownian motion plus α-stable Lévy process with $\alpha = 1.2$ (Example 5.2)

n	h	v	σ	pct	σ	pct
1000	5 min	0.000001	0	1	0.2	1
1000	5 min	0.0001	0	1	0.2	1
1000	5 min	0.01	0	1	0.2	1
1000	5 min	0.1	0	1	0.2	1
1000	1 min	0.0001	0	1	0.2	1
1000	1 hour	0.0001	0	1	0.2	1
1000	1 day	0.0001	0	1	0.2	1
100	5 min	0.0001	0	0.994	0.2	1
10 000	5 min	0.0001	0	1	0.2	1

where

$$\sigma_t = \mathrm{e}^{K_t}, \qquad \mathrm{d}K_t = -k(K_t - \bar{K})\,\mathrm{d}t + \varsigma\,\mathrm{d}W_t^{(2)}, \qquad d\langle W^{(1)}, W^{(2)}\rangle_t = \rho\,\mathrm{d}t, \tag{16}$$

$W^{(\ell)}$ are standard Brownian motions, $\ell = 1, 2, 3$, and $L_t = cG_t + \eta W^{(3)}_{G_t}$ is an independent variance gamma process, a pure-jump Lévy process with BG index $\alpha = 0$ [18]; G is a gamma subordinator independent of $W^{(3)}$ with $G_h \sim \Gamma(h/b, b)$. For σ, we choose $K_0 = \ln(0.3)$, $k = 0.09$, $\bar{K} = \ln(0.25)$, $\varsigma = 0.05$ to ensure that σ fluctuates in the range 0.2–0.4. As for the jump part of X, we use $\mathrm{Var}(G_1) = b = 0.23$, $\eta = 0.2$, $c = -0.2$, estimated from the S&P 500 index in [18]. The remaining parameters are $\rho = -0.7$ and $\mu = 0$. The following results in Table 5 confirm the reliability of the test for the presence of a Brownian component with $n = 1000$, $h = 5$ minutes and $v = 10^{-4}$.

Remark. In [21], a variable threshold function is used to estimate the volatility, in order to account for heteroscedasticity and volatility clustering, with results very similar to the ones ob-

Table 5. Testing for the presence of a Brownian component: stochastic volatility process with variance gamma jumps (Example 5.3)

n	h	v	σ	pct	σ	pct
1000	5 min	0.000001	0	0.032	Stoch.	1
1000	5 min	0.0001	0	0.017	Stoch.	1
1000	5 min	0.01	0	0.027	Stoch.	1
1000	5 min	0.1	0	0.054	Stoch.	1
1000	1 min	0.0001	0	0.034	Stoch.	1
1000	1 hour	0.0001	0	0.918	Stoch.	1
1000	1 day	0.0001	0	1.000	Stoch.	1
100	5 min	0.0001	0	0.049	Stoch.	1
10 000	5 min	0.0001	0	0.912	Stoch.	1

tained with a constant threshold. This is justified by the fact that in most applications, values of σ are within the range [0.1, 0.8], thus the order of magnitude of Λ in (7) is o 1.

6. Applications to financial time series

We apply our tests to explore the fine structure of price fluctuations in two financial time series. We consider the DM/USD exchange rate from October 1st, 1991 to November 29th, 1994 and the SPX futures prices from January 3rd, 1994 to December 18th, 1997. From high-frequency time series, we build five-minute log-returns (excluding, in the case of SPX futures, overnight log-returns). This sampling frequency avoids many microstructure effects seen at shorter time scales (e.g., seconds), while leaving us with a relatively large sample.

6.1. Deutsche Mark/USD exchange rate

The DM/USD exchange rate time series was compiled by Olsen & Associates. We consider the series of 64 284 equally spaced five-minute log-returns, with $h = \frac{1}{252\times 84} \approx 4.7\times 10^{-5}$, displayed in Figure 1.

Barndorff-Nielsen and Shephard [6] provide evidence for the presence of jumps in this series using nonparametric methods. Using as threshold $r_h = h^{0.999}$, we apply the test of Section 5.1 to the degree of activity of the jump component. As in the simulation study, we divide the data into 64 non-overlapping batches of $n = 1000$ observations each and compute, for each batch, the statistic $U_{h(j)}^{(\alpha)}$, $j = 1, \dots, 64$, with $v = 10^{-4}$. Only 4.7% of the values observed are outside the interval $[-1.96, 1.96]$, hence we cannot reject the assumption (H_0) $\alpha < 1$. Given this result, we can now use the test in Section 5.2 for the presence of a Brownian component in the price process. Computation of the statistic U_h shows values much larger than 1.96 for all batches: we reject (H_0) $\sigma \equiv 0$. These results indicate, for instance, that a variance gamma model, with no Brownian component, would be inadequate for the DM/USD time series.

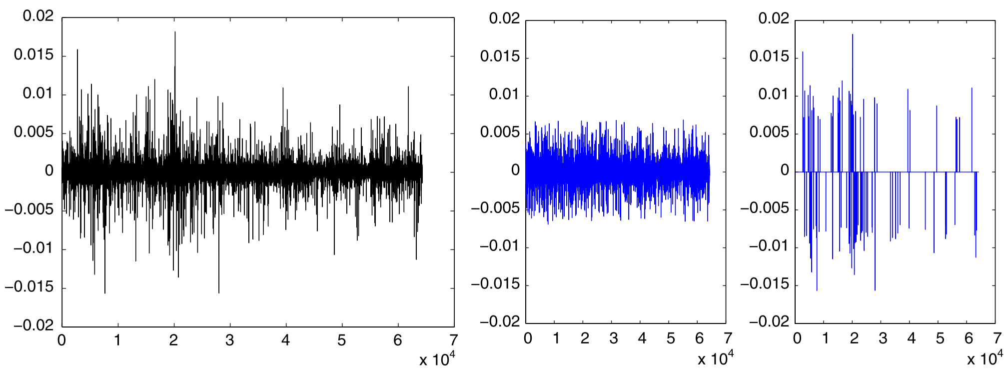

Figure 1. Left: DM/USD five-minute log-returns, October 1991 to November 1994. Center: plot of $\Delta_i X I_{\{(\Delta_i X)^2 \le r_h\}}$, $i = 1, \dots, n$. Right: increments with jumps $\Delta_i X I_{\{(\Delta_i X)^2 > r_h\}}$, $i = 1, \dots, n$.

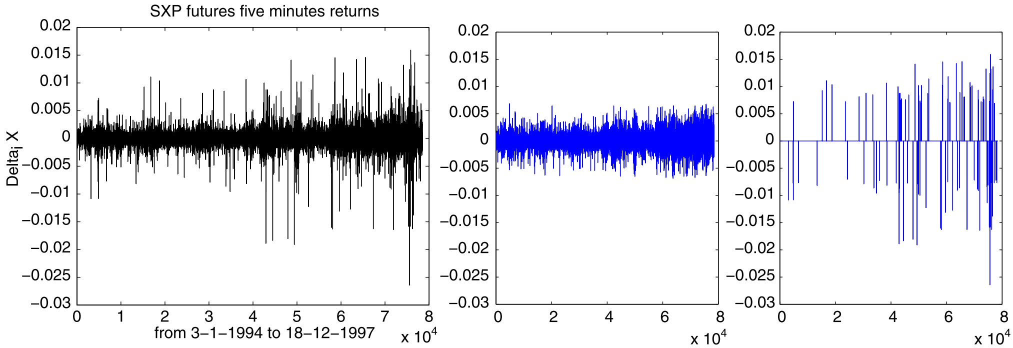

Figure 2. Left: SPX five-minute log-returns, January 1994 to December 1997. Center: plot of $\Delta_i X I_{\{(\Delta_i X)^2 \le r_h\}}$, $i = 1, \ldots, n$. Right: increments with jumps $\Delta_i X I_{\{(\Delta_i X)^2 > r_h\}}$, $i = 1, \ldots, n$.

6.2. S&P 500 index

We consider a series of 78 497 non-overlapping five-minute log-returns, as displayed in Figure 2. Using as threshold $r_h = h^{0.999}$, we decompose the series into periods displaying jumps and other periods, as displayed in Figure 2 (central and right panels).

We divide the data into 78 non-overlapping batches of $n = 1000$ observations each and compute, for each batch, the statistic $U^{(\alpha)}_{h(j)}$, $j = 1, \ldots, 64$, with $v = 10^{-4}$. 5.1% of the values observed are outside the interval $[-1.96, 1.96]$: for this period, we cannot reject the assumption (H_0) $\alpha < 1$. Given this result, we can use the test for the presence of a Brownian component in the price process. Computation of the statistic U_h shows values much larger than 1.96 for all batches: we reject (H_0) $\sigma \equiv 0$. The test thus indicates the presence of a Brownian martingale component.

We note that our findings contradict the conclusion of Carr *et al.* [8] who model the (log-) SPX index from 1994 to 1998 as a tempered stable Lévy process plus a Brownian motion and propose a pure-jump model using a parametric estimation method. Under less restrictive assumptions on the structure of the process and using our nonparametric test, we find evidence for a non-zero Brownian component in the index.

Appendix: Technical results and proofs

Proof of Lemma 3.1. By [23], Theorem 25.1, there exists a sequence (n_k) such that

$$\sup_{t_j \in \Pi^{(n_k)}} \left| (\Delta_j M)^2 - \sum_{s \in]t_{j-1}, t_j]} (\Delta M_s)^2 \right| \overset{\text{a.s.}}{\to} 0, \tag{17}$$

where $\Pi^{(n_k)}$ is the partition of $[0,T]$ on which the increments $(\Delta_i M)^2$ are constructed. Let us rename n_k as n. Using Itô's formula, we have

$$(\Delta_i M)^2 - \sum_{s\in]t_{i-1},t_i]} (\Delta M_s)^2 = 2\int_{t_{i-1}}^{t_i} (M_{s_-} - M_{t_{i-1}})\,\mathrm{d}M_s.$$

(i) For $\alpha < 1$, our statement is proved in [21], Lemma A.2, which uses the fact that the speed of convergence to 0 of $\sum_{i=1}^n |\int_{t_{i-1}}^{t_i}(M_{s_-} - M_{t_{i-1}})\,\mathrm{d}M_s|$ is shown in [12] to be $u_n = n$. For $\alpha = 1$, the same reasoning can be repeated since $u_n = n/(\log n)^2$ does not change the conclusion.

(ii) If $\alpha > 1$, we have $u_n = (n/\log n)^{1/\alpha}$ and can only conclude that a.s. for small h,

$$\sup_i \left|\int_{t_{i-1}}^{t_i} (M_{s_-} - M_{t_{i-1}})\,\mathrm{d}M_s\right| \le c u_n^{-1}$$

with $c > 0$, so that a.s. for small h, we have

$$\begin{aligned}\sup_i \Bigl(\sum_{s\in]t_{i-1},t_i]} (\Delta M_s)^2\Bigr) I_{\{(\Delta_i M)^2 \le 4r_h\}} &\le \sup_i \left|(\Delta_i M)^2 - \sum_{s\in]t_{i-1},t_i]} (\Delta M_s)^2\right| + \sup_i |(\Delta_i M)^2| \\ &\le c u_n^{-1} + 4r_h = \mathrm{O}\biggl(\delta^{1/\alpha} \log^{1/\alpha}\frac{1}{h}\biggr).\end{aligned}$$

□

Lemma A.1. *Under* (5):

(i) *there exists a strictly positive variable* $\bar{h}$ *such that for all* $i = 1,\dots,n$,

$$I_{\{h\le\bar{h}\}} I_{\{(\Delta_i X_0)^2 > r_h\}} = 0 \qquad a.s.; \tag{18}$$

(ii)

$$\forall c > 0,\, nP\{\Delta_i N \ne 0, (\Delta_i M)^2 > c r_h\} \xrightarrow{h\to 0} 0; \tag{19}$$

(iii) *in the case* $r_h = h^\beta$, $\beta \in]0,1[$, *we have*

$$\limsup_{h\to 0} h^{\alpha\beta/2} \sum_{i=1}^n P\{(\Delta_i X)^2 > r_h\} \le c. \tag{20}$$

Proof. Equality (18) is a consequence of (7), while (19) is a consequence of the independence of N and M, and of the Chebyshev inequality: as $h \to 0$,

$$nP\{\Delta_i N \ne 0, (\Delta_i M)^2 > c r_h\} \le n\mathrm{O}(h) \cdot \frac{E[(\Delta_i M)^2]}{c r_h} = \mathrm{O}\biggl(\frac{h}{r_h}\biggr).$$

The proof of (20) can be achieved as in [3], Lemma 6, but we give a simpler proof under our assumptions. It is sufficient to show that

$$P\{(\Delta_i X)^2 > r_h\} \le c h^{1-\alpha\beta/2}. \tag{21}$$

First, we show that

$$P\left\{|\Delta_i X| > \sqrt{r_h}\right\} = P\left\{|\Delta_i M| > \sqrt{r_h}/4\right\} + \mathrm{O}(h^{1-\alpha\beta/2}) \tag{22}$$

so that for (21), it is sufficient to prove that

$$P\left\{|\Delta_i M| > \sqrt{r_h}/4\right\} \leq c h^{1-\alpha\beta/2}. \tag{23}$$

To show (22), note that if $|\Delta_i X| > \sqrt{r_h}$, then either $\Delta_i J \neq 0$ or $|\Delta_i M| > \sqrt{r_h}/4$ since, for small h,

$$\sqrt{r_h} < |\Delta_i X| \leq |\Delta_i X_0| + |\Delta_i J| + |\Delta_i M| \leq \sqrt{r_h}/2 + |\Delta_i J| + |\Delta_i M| \qquad \text{a.s.} \tag{24}$$

Thus,

$$P\left\{|\Delta_i X| > \sqrt{r_h}\right\} \leq P\{\Delta_i J \neq 0\} + P\left\{|\Delta_i M| > \sqrt{r_h}/4\right\}$$

and since $P\{\Delta_i J \neq 0\} = \mathrm{O}(h) = \mathrm{o}(h^{1-\alpha\beta/2})$, (22) is verified.

In order to verify (23), define $\tilde{N}_t := \sum_{s \leq t} I_{\{|\Delta M_s| > \sqrt{r_h}/4\}}$ and write

$$\begin{aligned} P\left\{|\Delta_i M| > \sqrt{r_h}/4\right\} &= P\left\{\Delta_i \tilde{N} = 0, |\Delta_i M| > \sqrt{r_h}/4\right\} \\ &\quad + P\left\{\Delta_i \tilde{N} \geq 1, |\Delta_i M| > \sqrt{r_h}/4\right\} \\ &\leq P\{\Delta_i \tilde{N} \geq 1\} + P\left\{\Delta_i \tilde{N} = 0, |\Delta_i M| > \sqrt{r_h}/4\right\}. \end{aligned} \tag{25}$$

Note that $\tilde{N}_t = \int_0^t \int_{|x| > \sqrt{r_h}/4} \mu(\mathrm{d}x, \mathrm{d}t)$ is a compound Poisson process with intensity $\nu\{|x| > \sqrt{r_h}/4\} = \mathrm{O}(r_h^{-\alpha/2})$, so $P\{\Delta_i \tilde{N} \geq 1\} = \mathrm{O}(h\nu\{|x| > \sqrt{r_h}/4\}) = \mathrm{O}(h^{1-\alpha\beta/2})$ and thus the first term above is dominated by $h^{1-\alpha\beta/2}$, as required. Finally, on $\{\Delta_i \tilde{N} = 0\}$, M does not have jumps bigger than $\sqrt{r_h}/4$ on the interval $]t_{i-1}, t_i]$, so

$$\Delta_i M = \int_{t_{i-1}}^{t_i} \int_{|x| \leq \sqrt{r_h}/4} x \tilde{\mu}(\mathrm{d}x, \mathrm{d}t) - h \int_{\sqrt{r_h}/4 < |x| \leq 1} x \nu(\mathrm{d}x),$$

therefore

$$\begin{aligned} P\left\{\Delta_i \tilde{N} = 0, |\Delta_i M| > \sqrt{r_h}/4\right\} &\leq P\left\{|\Delta_i M| > \sqrt{r_h}/4, |\Delta M_s| \leq \sqrt{r_h}/4 \text{ for all } s \in]t_{i-1}, t_i]\right\} \\ &\leq 4 \frac{E[(\Delta_i M)^2 I_{\{|\Delta M_s| \leq \sqrt{r_h}/4 \text{ for all } s \in]t_{i-1}, t_i]\}}]}{r_h} \\ &= \mathrm{O}\left(\frac{h\eta^2(r_h/4)}{r_h}\right) = \mathrm{O}(h^{1-\alpha\beta/2}) \end{aligned}$$

and (23) is verified. □

Proof of Proposition 3.3.

$$\begin{aligned}
&\frac{\sum_i (\Delta_i X)^4 I_{\{(\Delta_i X)^2 \le r_h\}}}{3h} \\
&\quad = \frac{\sum_i (\Delta_i X_1)^4 I_{\{(\Delta_i X_1)^2 \le 4r_h\}}}{3h} + \frac{1}{3h} \sum_i (\Delta_i X_1)^4 \big(I_{\{(\Delta_i X)^2 \le r_h\}} - I_{\{(\Delta_i X_1)^2 \le 4r_h\}} \big) \\
&\qquad + \sum_{k=1}^{4} \binom{4}{k} \frac{\sum_i (\Delta_i X_1)^{4-k} (\Delta_i M)^k I_{\{(\Delta_i X)^2 \le r_h\}}}{3h} \\
&\quad := \sum_{j=1}^{3} I_j(h).
\end{aligned}$$

By Proposition 1 in [20], $I_1(h)$ tends to $\int_0^T \sigma_t^4 \,\mathrm{d}t$ in probability. We show here that the other terms tend to zero in probability. Let us consider $I_2(h) := \frac{1}{3h} \sum_i (\Delta_i X_1)^4 (I_{\{(\Delta_i X)^2 \le r_h\}} - I_{\{(\Delta_i X_1)^2 \le 4r_h\}})$: on $\{(\Delta_i X)^2 \le r_h, (\Delta_i X_1)^2 > 4r_h\}$, we have

$$\sqrt{r_h} \ge |\Delta_i X| > |\Delta_i X_1| - |\Delta_i M| > 2\sqrt{r_h} - |\Delta_i M|, \tag{26}$$

so $|\Delta_i M| > \sqrt{r_h}$. Moreover, if $|\Delta_i X_1| > 2\sqrt{r_h}$, then we necessarily have $\Delta_i N \neq 0$ since

$$|\Delta_i X_0| + |\Delta_i J| \ge |\Delta_i X_1| > 2\sqrt{r_h} \tag{27}$$

and, by (18), a.s. for sufficiently small h, for all $i = 1, \dots, n$, $|\Delta_i X_0| \le \sqrt{r_h}$, thus $|\Delta_i J| > 2\sqrt{r_h} - |\Delta_i X_0| \ge \sqrt{r_h}$. It follows that

$$P\left\{ \frac{1}{h} \sum_i (\Delta_i X_1)^4 I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i X_1)^2 > 4r_h\}} \neq 0 \right\} \le n P\big\{ |\Delta_i M| > \sqrt{r_h}, \Delta_i N \neq 0 \big\} \to 0,$$

by Lemma A.1. On the other hand, for all $i = 1, \dots, n$ on $\{(\Delta_i X_1)^2 \le 4r_h\}$, we have, for sufficiently small h, $\Delta_i N = 0$ because

$$|\Delta_i J| - |\Delta_i X_0| \le |\Delta_i X_1| \le 2\sqrt{r_h}, \tag{28}$$

so if $\Delta_i N \neq 0$, then a.s. for small h, we in fact have $\Delta_i N = 1$ and $\Delta J_s \ge 1$, by the definition of J. Therefore, if $\Delta_i N \neq 0$, we would have $1 \le |\Delta_i J| \le 2\sqrt{r_h} + \sqrt{r_h} = 3\sqrt{r_h}$, which is impossible for small h. It follows that

$$\begin{aligned}
\{(\Delta_i X)^2 > r_h, (\Delta_i X_1)^2 \le 4r_h\} &\subset \{(\Delta_i X_0 + \Delta_i M)^2 > r_h\} \\
&\subset \left\{ (\Delta_i X_0)^2 > \frac{r_h}{4} \right\} \cup \left\{ (\Delta_i M)^2 > \frac{r_h}{4} \right\}.
\end{aligned}$$

This implies, by (18) and (23), that a.s. as $h \to 0$,

$$\frac{1}{h}\sum_i (\Delta_i X_1)^4 I_{\{(\Delta_i X)^2 > r_h, (\Delta_i X_1)^2 \le 4r_h\}} \le \frac{\sum_i (\Delta_i X_0)^4 I_{\{(\Delta_i M)^2 > r_h/4\}}}{h}$$
$$\le \Lambda^4 h \ln^2 \frac{1}{h}\sum_i I_{\{(\Delta_i M)^2 > r_h/4\}} \xrightarrow{P} 0.$$

We can conclude that $I_2(h) \xrightarrow{P} 0$ as $h \to 0$. Now, consider $I_3(h) := \sum_{k=1}^4 \binom{4}{k} I_{3,k}(h)$, where

$$I_{3,k}(h) := \frac{1}{3h}\sum_i (\Delta_i X_1)^{4-k}(\Delta_i M)^k I_{\{(\Delta_i X)^2 \le r_h\}}, \qquad k = 1, \dots, 4,$$

is decomposable as

$$\begin{aligned}&\frac{1}{3h}\sum_i (\Delta_i X_1)^{4-k}(\Delta_i M)^k I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 \le 4r_h\}}\\ &\quad + \frac{1}{3h}\sum_i (\Delta_i X_1)^{4-k}(\Delta_i M)^k I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 > 4r_h\}}.\end{aligned} \tag{29}$$

We have, a.s. for small h, that for all i on $\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 > 4r_h\}$, $\Delta_i N \ne 0$ since

$$2\sqrt{r_h} - |\Delta_i X_1| < |\Delta_i M| - |\Delta_i X_1| \le |\Delta_i X| \le \sqrt{r_h}$$

and then $|\Delta_i X_1| > \sqrt{r_h}$ and, similarly as in (27), $|\Delta_i J| > 3\sqrt{r_h}/4$. So, the probability that the second term of (29) differs from zero is bounded by (19) and tends to zero. As for the first term, a.s. for sufficiently small h, for all i on $\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 \le 4r_h\}$, we have $\Delta_i N = 0$ because

$$|\Delta_i X_1| - |\Delta_i M| \le |\Delta_i X| \le \sqrt{r_h},$$

thus $|\Delta_i X_1| < 3\sqrt{r_h}$ and we proceed as in (28). So, the first term in (29) is a.s. dominated by

$$\frac{\sum_i |\Delta_i X_1|^{4-k}|\Delta_i M|^k I_{\{\Delta_i N = 0, (\Delta_i M)^2 \le 4r_h\}}}{3h} \le \frac{\sum_i |\Delta_i X_0|^{4-k}|\Delta_i M|^k I_{\{(\Delta_i M)^2 \le 4r_h\}}}{3h}.$$

Now, for $k = 4$, we apply to M property (C.19) in [4], Lemma 5, with β there being α here, $u_n = \sqrt{r_h} = h^{\beta/2}$, $p = 4$, $v_h = h^\phi$ for a proper exponent ϕ we specify below and $\beta' = 0$. Result (C.19) of [4] then implies that

$$\frac{1}{h}E\left[\left|\sum_{i=1}^n (\Delta_i M)^4 I_{\{|\Delta_i M| \le 2\sqrt{r_h}\}} - \sum_{v \le T} |\Delta M_v|^4 I_{\{|\Delta M_v| \le 2\sqrt{r_h}\}}\right|\right] \le c h^{(\beta/2)(4-\alpha)-1} \cdot \eta_{4,n},$$

where $\eta_{4,n} = h(h^{\beta/2}v_h)^{-\alpha} + h^2 h^{\alpha\beta/2}(h^{\beta/2}v_h)^{-3\alpha} + h h^{\alpha\beta/2}(h^{\beta/2})^{-2\alpha} + (2h^{\beta/2})^\alpha + h^{1/4}h^{-((4-\alpha)/4)\beta/2} + v_h^{(4-\alpha)/4}$. As soon as $\beta > 1/(2-\alpha/2)$ and we choose $\phi \in]0, \frac{1-\beta}{3}[$, so that for all $\alpha \in]0,2[$ we have $\phi < (2/\alpha - \beta)/3$, it is guaranteed both that $h^{(\beta/2)(4-\alpha)-1} \to 0$ and that $h^{(\beta/2)(4-\alpha)-1} \cdot \eta_{4,n} \to 0$. Thus,

$$\lim_h \frac{\sum_i |\Delta_i M|^4 I_{\{(\Delta_i M)^2 \le 4r_h\}}}{3h} = \lim_h \frac{\sum_i \int_{t_{i-1}}^{t_i} \int_{|x|\le 2\sqrt{r_h}} |x|^4 \mu(\mathrm{d}x, \mathrm{d}t)}{3h}$$

and

$$\begin{aligned} E\biggl[\sum_i \int_{t_{i-1}}^{t_i} \int_{|x|\le 2\sqrt{r_h}} |x|^4 \mu(\mathrm{d}x,\mathrm{d}t)/3h\biggr] &= \mathrm{O}\biggl(\int_{|x|\le 2\sqrt{r_h}} |x|^4 \nu(\mathrm{d}x)/h\biggr) \\ &= \mathrm{O}\bigl(h^{(\beta/2)(4-\alpha)-1}\bigr) \to 0, \end{aligned}$$

given that $\beta > 1/(2-\alpha/2)$.

To show, further, that the terms

$$\frac{\sum_i |\Delta_i X_0|^{4-k} |\Delta_i M|^k I_{\{(\Delta_i M)^2 \le 4r_h\}}}{3h}$$

tend to zero in probability for $k = 1,2,3,$ we use the fact that, by (11), each term is dominated by (recall the notation in (10))

$$c\frac{\sum_i |\Delta_i X_0|^{4-k} |\Delta_i M^{(h)}|^k}{3h} + c\frac{\sum_i |\Delta_i X_0|^{4-k} |hd(2\sqrt[4]{r_h})|^k}{3h}.$$

Now, a.s.

$$\begin{aligned} &\frac{\sum_i |\Delta_i X_0|^{4-k} |hd(2\sqrt[4]{r_h})|^k}{3h} \\ &\quad \le \left(h \ln\frac{1}{h}\right)^{(4-k)/2} nh^{k-1} \biggl[\bigl|c + r_h^{(1-\alpha)/4}\bigr|^k I_{\{\alpha\ne 1\}} + \ln^k \frac{1}{r_h^{1/4}} I_{\{\alpha=1\}}\biggr] \\ &\quad \le ch^{k/2}\left(\ln\frac{1}{h}\right)^{(4-k)/2} + ch^{k/2}\left(\ln\frac{1}{h}\right)^{(4-k)/2} r_h^{k(1-\alpha)/4} + h^{h/2} \ln^{2-k/2} \frac{1}{hr_h^{1/4}} \\ &\quad = \mathrm{o}(1) + ch^{k[1/2+\beta(1-\alpha)/4]} \log^{(4-k)/2} \frac{1}{h} \to 0 \end{aligned}$$

for all $k = 1,2,3$ as $r_h = h^\beta$, $\beta \in]0,1[$. As for

$$\frac{\sum_i |\Delta_i X_0|^{4-k} |\Delta_i M^{(h)}|^k}{3h}, \tag{30}$$

we need to deal separately with each of $k = 1,2,3$. Note that since a and σ are locally bounded on $\Omega \times [0,T]$, we can assume that they are bounded without loss of generality, so

$E[(\int_{t_{i-1}}^{t_i} \sigma_s \,\mathrm{d}W_s)^{2k}] = \mathrm{O}(h^k)$ for each $k = 1, 2, 3$, using, for instance, the Burkholder inequality [24], page 226, and a.s. $(\int_{t_{i-1}}^{t_i} a_s \,\mathrm{d}s)^{2k} = \mathrm{o}(h^k)$. Therefore, $E[(\Delta_i X_0)^{2k}] = \mathrm{O}(h^k)$ for each of $k = 1, 2, 3$. For $k = 1$, the expected value of (30) is bounded by $(n/3h)\sqrt{E[(\Delta_i X_0)^6]} \times \sqrt{E(\Delta_i M^{(h)})^2} = \mathrm{O}(r_h^{(1/4)(1-\alpha/2)})$ and thus tends to zero as $h \to 0$. As for $k = 2$,

$$\frac{\sum_i (\Delta_i X_0)^2 (\Delta_i M^{(h)})^2}{h} \le h \ln \frac{1}{h} \frac{\sum_i (\Delta_i M^{(h)})^2}{h}, \tag{31}$$

whose expected value is given by

$$\ln \frac{1}{h} \eta^2 (2 r_h^{1/4}) \to 0$$

as $h \to 0$ since $r_h = h^\beta$, with $\beta > 0$. Concerning $k = 3$, we have

$$\frac{\sum_i |\Delta_i X_0| |\Delta_i M^{(h)}|^3}{h} \le \frac{c}{h} \sum_i (\Delta_i X_0)^2 \bigl(\Delta_i M^{(h)}\bigr)^2 + \frac{c}{h} \sum_i \bigl(\Delta_i M^{(h)}\bigr)^4,$$

so that this step is reduced to the steps with $k = 2, 4$ which we dealt with previously. $\square$

Proof of Theorem 3.4. Let us define $K_{ni} := (\int_{t_{i-1}}^{t_i} \int_{|x| \le \varepsilon} x \tilde{\mu}(\mathrm{d}x, \mathrm{d}t) - h \int_{\varepsilon < |x| \le 1} x \nu(\mathrm{d}x))^2$. We apply the Lindeberg–Feller theorem to the double array sequence H_{ni} given by the normalized versions of the variables K_{ni}, $i = 1, \ldots, n$, and $n = T/h$. Using relations (9), we have

$$\begin{aligned} E[K_{ni}] &= h \ell_{2,h} \varepsilon^{2-\alpha} + \left(h \int_{\varepsilon < |x| \le 1} x \nu(\mathrm{d}x) \right)^2 \\ &= h \ell_{2,h} \varepsilon^{2-\alpha} + \ell_{1,h}^2 h^2 \left[(c + \varepsilon^{1-\alpha})^2 I_{\{\alpha \ne 1\}} + \left(\ln^2 \frac{1}{\varepsilon} \right) I_{\{\alpha = 1\}} \right]. \end{aligned} \tag{32}$$

Taking $\varepsilon = h^u$ and any $u \in]0, 1/2]$, we obtain that

$$\begin{aligned} v_{ni}^2 := \mathrm{var}[K_{ni}] = E\left[\left(\int_{t_{i-1}}^{t_i} \int_{|x| \le \varepsilon} x \tilde{\mu}(\mathrm{d}x, \mathrm{d}t) - h \int_{\varepsilon < |x| \le 1} x \nu(\mathrm{d}x) \right)^4 \right] \\ - E_{ni}^2 \sim h \int_{|x| \le \varepsilon} x^4 \nu(\mathrm{d}x) = h \ell_{4,h} \varepsilon^{4-\alpha} \end{aligned}$$

as $h \to 0$. Then, consider

$$H_{ni} := \frac{K_{ni} - E[K_{ni}]}{\sqrt{n} v_{ni}} \sim \frac{K_{ni} - h \ell_{2,h} \varepsilon^{2-\alpha} - \ell_{1,h}^2 h^2 [(c + \varepsilon^{1-\alpha})^2 I_{\{\alpha \ne 1\}} + (\ln^2 1/\varepsilon) I_{\{\alpha = 1\}}]}{\sqrt{T} \sqrt{\ell_{4,h}} \varepsilon^{2-\alpha/2}}.$$

We now show that for any $\delta > 0$, there exists a $q > 1$ such that

$$\sum_{i=1}^n E\bigl[H_{ni}^2 I_{\{|H_{ni}| > \delta\}}\bigr] \le c \varepsilon^{\alpha/(2q)} \to 0 \tag{33}$$

as $h \to 0$, so the Lindeberg condition is satisfied and implies that

$$\sum_{i=1}^{n} H_{ni} \xrightarrow{d} N(0,1). \tag{34}$$

Noting that $h/\varepsilon^{2-\alpha/2}$ and $(h\varepsilon^{1-\alpha})/(\varepsilon^{2-\alpha/2})I_{\{\alpha\neq 1\}} + (h\ln^2(1/\varepsilon))/(\varepsilon^{2-\alpha/2})I_{\{\alpha=1\}}$ tend to zero as $h \to 0$, (34) leads to (12). To show inequality (33), consider

$$nE\big[H_{n1}^2 I_{\{|H_{n1}|>\delta\}}\big] \le nE^{1/p}[H_{n1}^{2p}]P^{1/q}\{|H_{n1}|>\delta\}, \tag{35}$$

as for the last factor above, we note that $|H_{n1}| > \delta$ if and only if either

$$\begin{aligned} K_{n1} &< h\ell_{2,h}\varepsilon^{2-\alpha} + \ell_{1,h}^2 h^2\left[(c+\varepsilon^{1-\alpha})^2 I_{\{\alpha\neq 1\}} + \left(\ln^2\frac{1}{\varepsilon}\right)I_{\{\alpha=1\}}\right] - \delta\sqrt{T\ell_{4,h}}\varepsilon^{2-\alpha/2} \\ &= \varepsilon^{2-\alpha/2}\big(\mathrm{o}(1) - c\delta\big), \end{aligned}$$

where c denotes a generic constant, or

$$K_{n1} > h\ell_{2,h}\varepsilon^{2-\alpha} + \ell_{1,h}^2 h^2\left[(c+\varepsilon^{1-\alpha})^2 I_{\{\alpha\neq 1\}} + \left(\ln^2\frac{1}{\varepsilon}\right)I_{\{\alpha=1\}}\right] + c\delta\varepsilon^{2-\alpha/2} = \mathrm{O}(\varepsilon^{2-\alpha/2}).$$

However, $K_{n1} \ge 0$, while for sufficiently small h, the right-hand term of the first inequality above is strictly negative, therefore $|H_{n1}| > \delta$ if and only if $K_{n1} > c\varepsilon^{2-\alpha/2}$, that is, either

$$-c\varepsilon^{1-\alpha/4} \sim h(c+\varepsilon^{1-\alpha})I_{\{\alpha\neq 1\}} + I_{\{\alpha=1\}}h\ln\frac{1}{\varepsilon} - c\varepsilon^{1-\alpha/4} > \int_0^{t_1}\int_{|x|\le\varepsilon} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t)$$

or, for sufficiently small h, $\int_0^{t_1}\int_{|x|\le\varepsilon} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t) > c\varepsilon^{1-\alpha/4}$, and so $|H_{n1}| > \delta$ if and only if

$$\left|\int_0^{t_1}\int_{|x|\le\varepsilon} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t)\right| > c\varepsilon^{1-\alpha/4}.$$

This entails that for sufficiently small h,

$$\begin{aligned} P\{|H_{n1}|>\delta\} &= P\left\{\left|\int_0^{t_1}\int_{|x|\le\varepsilon} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t)\right| > c\varepsilon^{1-\alpha/4}\right\} \\ &\le c\frac{E[|\int_0^{t_1}\int_{|x|\le\varepsilon} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t)|^2]}{\varepsilon^{2-\alpha/2}} = h^{1-(\alpha u)/2} \to 0. \end{aligned}$$

The first two factors of the right-hand side of (35) are dominated by

$$\begin{aligned} &cn\frac{E^{1/p}[(K_{n1} - h\ell_{2,h}\varepsilon^{2-\alpha} - h^2\ell_{1,h}^2[(c+\varepsilon^{1-\alpha})^2 I_{\{\alpha\neq 1\}} + (\ln^2 1/\varepsilon)I_{\{\alpha=1\}}])^{2p}]}{\varepsilon^{4-\alpha}} \\ &\quad\le cn\frac{E^{1/p}[K_{n1}^{2p}] + (h\varepsilon^{2-\alpha})^2 + h^4(1-\varepsilon^{1-\alpha})^4 + h^4\ln^4 1/\varepsilon}{\varepsilon^{4-\alpha}}. \end{aligned}$$

The last three terms give no contribution to (35) since

$$n\frac{(h\varepsilon^{2-\alpha})^2+h^4(1-\varepsilon^{1-\alpha})^4+h^4\ln^4 1/\varepsilon}{\varepsilon^{4-\alpha}}h^{(1-\alpha u/2)(1/q)}\to 0.$$

On the other hand, by choosing, for example, $p=5/4$, we have

$$E[K_{n1}^{2p}]=\mathrm{O}(h\varepsilon^{5-\alpha}),$$

so we are left to deal with $n\frac{(h\varepsilon^{5-\alpha})^{1/p}}{\varepsilon^{4-\alpha}}h^{(1-\alpha u/2)(1/q)}=\varepsilon^{\alpha/(2q)}$ and the inequality in (33) is proved. □

Lemma A.2. *As $h\to 0$, if $r_h\to 0$, $n=T/h$ and $\sup_{i=1,\dots,n}|a_{hi}|=\mathrm{O}(r_h)$, then*

$$\sum_i |a_{hi}|I_{\{(\Delta_i X)^2\le r_h\}}-\sum_i |a_{hi}|I_{\{(\Delta_i M)^2\le 4r_h,\Delta_i N=0\}}\stackrel{P}{\to}0.$$

Proof. On $\{(\Delta_i X)^2\le r_h\}$, we have $|\Delta_i L|-|\Delta_i X_0|\le|\Delta_i X|\le\sqrt{r_h}$ and, thus, by (7), for small h, $|\Delta_i L|\le 2\sqrt{r_h}$, so that a.s.

$$\lim_{h\to 0}\sum_i |a_{hi}|I_{\{(\Delta_i X)^2\le r_h\}}\le\lim_{h\to 0}\sum_i |a_{hi}|I_{\{(\Delta_i L)^2\le 4r_h\}}.$$

However,

$$\sum_i |a_{hi}|I_{\{(\Delta_i L)^2\le 4r_h,\Delta_i N\ne 0\}}\le\sup_i|a_{hi}|N_T\stackrel{a.s.}{\to}0 \tag{36}$$

as $h\to 0$ and thus a.s.

$$\lim_{h\to 0}\sum_i |a_{hi}|I_{\{(\Delta_i X)^2\le r_h\}}\le\lim_{h\to 0}\sum_i |a_{hi}|I_{\{(\Delta_i L)^2\le 4r_h,\Delta_i N=0\}}=\lim_{h\to 0}\sum_i |a_{hi}|I_{\{(\Delta_i M)^2\le 4r_h,\Delta_i N=0\}}.$$

We now show that, on the other hand, the positive quantity

$$\lim_{h\to 0}\sum_i |a_{hi}|\big(I_{\{(\Delta_i L)^2\le 4r_h,\Delta_i N=0\}}-I_{\{(\Delta_i X)^2\le r_h\}}\big)=0\qquad\text{a.s.}$$

In fact,

$$\begin{aligned}
&\{(\Delta_i L)^2\le 4r_h,\ \Delta_i N=0\}-\{(\Delta_i X)^2\le r_h\}\\
&\quad=\{(\Delta_i L)^2\le 4r_h,\ \Delta_i N=0,\ (\Delta_i X)^2>r_h\}\\
&\qquad\subset\big\{|\Delta_i L|\le 2\sqrt{r_h},\ \Delta_i N=0,\ |\Delta_i X_0|+|\Delta_i M|>\sqrt{r_h}\big\}\\
&\qquad\subset\big\{|\Delta_i X_0|>\sqrt{r_h}/2\big\}\cup\big\{|\Delta_i M|\le 2\sqrt{r_h},\ |\Delta_i M|>\sqrt{r_h}/2\big\}.
\end{aligned}$$

Since, by (18), a.s. for sufficiently small h $\sum_i |a_{hi}| I_{\{|\Delta_i X_0| > \sqrt{r_h}/2\}} = 0$, we a.s. have

$$\lim_{h\to 0} \sum_i |a_{hi}| \big(I_{\{(\Delta_i L)^2 \le 4r_h, \Delta_i N = 0\}} - I_{\{(\Delta_i X)^2 \le r_h\}}\big) \le \lim_{h\to 0} \sum_i |a_{hi}| I_{\{|\Delta_i M| \le 2\sqrt{r_h}, |\Delta_i M| > \sqrt{r_h}/2\}};$$

however, by Remark 3.2, as $h \to 0$,

$$\begin{aligned} E\bigg[\sum_i |a_{hi}| I_{\{|\Delta_i M| \le 2\sqrt{r_h}, |\Delta_i M| > \sqrt{r_h}/2\}}\bigg] &\le \mathrm{O}(r_h) n P\big\{|\Delta_i M| \le 2\sqrt{r_h}, |\Delta_i M| > \sqrt{r_h}/2\big\} \\ &\le \mathrm{O}(r_h) n P\big\{|\Delta_i M| I_{\{|\Delta_i M| \le 2\sqrt{r_h}\}} > \sqrt{r_h}/2\big\} \\ &\le \mathrm{O}(r_h) n \frac{E[(\Delta_i M)^2 I_{\{|\Delta_i M| \le 2\sqrt{r_h}\}}]}{r_h} \\ &= \mathrm{O}(r_h) n \frac{h\eta^2 (2 r_h^{c1/4})}{r_h} \to 0. \end{aligned}$$
□

Lemma A.3. *Under the assumptions of Theorem 3.5, for all $\alpha \in [0, 2[$,*

$$\begin{aligned} \sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i M)^2 \le r_h/16\}} - \mathrm{o}_P(h^{1-\alpha/2}) &\le \sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 \le 4r_h\}} \\ &\le \sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i M)^2 \le 9r_h/4\}} + \mathrm{o}_P(h^{1-\alpha/2}) \qquad a.s. \end{aligned} \tag{37}$$

Proof. Let us first deal with $\sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i X)^2 > r_h, (\Delta_i M)^2 \le 4r_h\}}$.

As in (24), on $\{(\Delta_i X)^2 > r_h\}$, we have either $|\Delta_i J| > \sqrt{r_h}/4$ or $|\Delta_i M| > \sqrt{r_h}/4$, so

$$\begin{aligned} &\sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i X)^2 > r_h, (\Delta_i M)^2 \le 4r_h\}} \\ &\quad \le \sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i X)^2 > r_h, \Delta_i J \ne 0, (\Delta_i M)^2 \le 4r_h\}} \\ &\qquad + \sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i X)^2 > r_h, (\Delta_i M)^2 > r_h/16, (\Delta_i M)^2 \le 4r_h\}}. \end{aligned}$$

However,

$$E\bigg[\frac{\sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i M)^2 \le 4r_h, \Delta_i N \ne 0\}}}{h^{1-\alpha/2}}\bigg] = \mathrm{O}\bigg(\frac{h\eta^2 (r_h^{1/4}) N_T}{h^{1-\alpha/2}}\bigg) \to 0,$$

so

$$\begin{aligned}
&\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i X)^2>r_h,(\Delta_i M)^2\le 4r_h\}}\\
&\quad\le o_P(h^{1-\alpha/2})+\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 4r_h,(\Delta_i M)^2>r_h/16\}}\\
&\quad= o_P(h^{1-\alpha/2})+\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 4r_h\}}-\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 4r_h,(\Delta_i M)^2\le r_h/16\}}\\
&\quad= o_P(h^{1-\alpha/2})+\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 4r_h\}}-\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le r_h/16\}}.
\end{aligned}\tag{38}$$

Now, consider $\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 4r_h,(\Delta_i M)^2>9r_h/4\}}$: on $\{2\sqrt{r_h}\ge|\Delta_i M|>\frac{3}{2}\sqrt{r_h}\}$, either $\Delta_i N\ne 0$, in which case

$$\frac{\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 4r_h,\Delta_i N\ne 0\}}}{h^{1-\alpha/2}}\xrightarrow{P}0,$$

as before, or $\Delta_i N=0$, in which case $|\Delta_i X|>|\Delta_i M|-|\Delta_i X_0|>\frac{3}{2}\sqrt{r_h}-\frac{1}{2}\sqrt{r_h}=\sqrt{r_h}$, so

$$\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i X)^2>r_h,(\Delta_i M)^2\le 4r_h\}}+o_P(h^{1-\alpha/2})\ge\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 4r_h,(\Delta_i M)^2>9r_h/4\}}.$$

Therefore,

$$\begin{aligned}
&\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i X)^2>r_h,(\Delta_i M)^2\le 4r_h\}}\\
&\quad\ge -o_P(h^{1-\alpha/2})+\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 4r_h\}}-\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 9r_h/4\}}.
\end{aligned}\tag{39}$$

Now combining (38) and (39), we obtain (37) since

$$\begin{aligned}
&\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i X)^2\le r_h,(\Delta_i M)^2\le 4r_h\}}\\
&\quad=\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 4r_h\}}-\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i X)^2>r_h,(\Delta_i M)^2\le 4r_h\}}.
\end{aligned}$$

□

Proof of Theorem 3.5. Note that under $\beta > \frac{1}{2-\alpha/2}$, the assumptions of Proposition 3.3 are satisfied. Since $X = X_1 + M$, we decompose

$$\frac{\hat{IV}_h - IV}{\sqrt{2h\hat{IQ}_h}} = \frac{\sum_{i=1}^n (\Delta_i X)^2 I_{\{(\Delta_i X)^2 \le r_h\}} - IV}{\sqrt{2h}\sqrt{\sum_i (\Delta_i X)^4 I_{\{(\Delta_i X)^2 \le r_h\}}/3h}} = \frac{\sum_{i=1}^n (\Delta_i X_1)^2 I_{\{(\Delta_i X_1)^2 \le 4r_h\}} - IV}{\sqrt{(2/3)\sum_i (\Delta_i X)^4 I_{\{(\Delta_i X)^2 \le r_h\}}}} \tag{40}$$

$$\begin{aligned} &+ \frac{\sqrt{2hIQ}}{\sqrt{(2/3)\sum_i (\Delta_i X)^4 I_{\{(\Delta_i X)^2 \le r_h\}}}} \\ &\times \biggl[\frac{\sum_{i=1}^n (\Delta_i X_1)^2 (I_{\{(\Delta_i X)^2 \le r_h\}} - I_{\{(\Delta_i X_1)^2 \le 4r_h\}})}{\sqrt{2hIQ}} \\ &\quad + 2\frac{\sum_{i=1}^n \Delta_i X_1 \Delta_i M I_{\{(\Delta_i X)^2 \le r_h\}}}{\sqrt{2hIQ}} + \frac{\sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i X)^2 \le r_h\}}}{\sqrt{2hIQ}} \biggr] \end{aligned} \tag{41}$$

$$:= \sum_{j=1}^4 I_j(h).$$

The proof of [20], Theorem 2, shows that $I_1(h)$ converges stably in law to a standard Gaussian random variable. To show that the remaining terms either tend to zero or to infinity, we can assume without loss of generality that both a and σ are bounded a.s. If $(\Delta_i X)^2 \le r_h$ and $(\Delta_i X_1)^2 > 4r_h$, then $|\Delta_i M| > \sqrt{r_h}$ and $\Delta_i N \ne 0$, exactly as for $I_2(h)$ in Proposition 3.3. It follows that

$$\begin{aligned} &P\biggl\{ \frac{\sum_{i=1}^n (\Delta_i X_1)^2 I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i X_1)^2 > 4r_h\}}}{\sqrt{2hIQ}} \ne 0 \biggr\} \\ &\quad \le nP\bigl\{\Delta_i N \ne 0, |\Delta_i M| > \sqrt{r_h}\bigr\} \to 0, \end{aligned}$$

by (19). The main factor of the remaining part of $I_2(h)$ is

$$\frac{\sum_{i=1}^n (\Delta_i X_1)^2 I_{\{(\Delta_i X)^2 > r_h, (\Delta_i X_1)^2 \le 4r_h\}}}{\sqrt{2hIQ}}.$$

We recall that on $\{|\Delta_i X_1| \le 2\sqrt{r_h}\}$, we have $\Delta_i N = 0$, thus $(\Delta_i X_1)^2 = (\Delta_i X_0)^2$. Moreover,

$$\frac{\sum_{i=1}^n (\int_{t_{i-1}}^{t_i} a_u \, \mathrm{d}u)^2 I_{\{(\Delta_i X)^2 > r_h, (\Delta_i X_1)^2 \le 4r_h\}}}{\sqrt{2hIQ}} = \mathrm{O}_P\bigl(\sqrt{h}\bigr) \to 0$$

and, by (20),

$$\frac{1}{\sqrt{2hIQ}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} a_u\,\mathrm{d}u\int_{t_{i-1}}^{t_i}\sigma_u\,\mathrm{d}W_u\,I_{\{(\Delta_i X)^2>r_h,(\Delta_i X_1)^2\le 4r_h\}}\le c\sqrt{h}\sqrt{h\ln\frac{1}{h}}\sum_{i=1}^{n}I_{\{(\Delta_i X)^2>r_h\}}$$

$$=\mathrm{O}\left(h^{1-\alpha\beta/2}\sqrt{\ln\frac{1}{h}}\right)\to 0.$$

Therefore, in probability,

$$\lim_{h\to 0}I_2(h)=\lim_{h\to 0}-\frac{\sum_{i=1}^{n}(\int_{t_{i-1}}^{t_i}\sigma_u\,\mathrm{d}W_u)^2 I_{\{(\Delta_i X)^2>r_h,(\Delta_i X_1)^2\le 4r_h\}}}{\sqrt{2hIQ}}.$$

We now show that term $I_3(h)/2$ in (41) tends to zero in probability. First, recall that $\Delta_i X_1=\Delta_i X_0+\Delta_i J$ and, within the sum $\sum_{i=1}^{n}\Delta_i J\Delta_i M I_{\{(\Delta_i X)^2\le r_h\}}/\sqrt{h}$, the term i contributes only when $\Delta_i N\ne 0$, in which case we also have $(\Delta_i X_1)^2>4r_h$ and thus $|\Delta_i M|>\sqrt{r_h}$, as in (26). That implies

$$P\left\{\frac{\sum_{i=1}^{n}\Delta_i J\Delta_i M I_{\{(\Delta_i X)^2\le r_h\}}}{\sqrt{2hIQ}}\ne 0\right\}\le nP\left\{\Delta_i N\ne 0,|\Delta_i M|>\sqrt{r_h}\right\}\to 0.$$

As for $\frac{\sum_{i=1}^{n}\Delta_i X_0\Delta_i M I_{\{(\Delta_i X)^2\le r_h\}}}{\sqrt{h}}$, as in the proof of Lemma A.2, we have

$$\frac{\sum_{i=1}^{n}\Delta_i X_0\Delta_i M I_{\{(\Delta_i X)^2\le r_h\}}}{\sqrt{h}}=\frac{\sum_{i=1}^{n}\Delta_i X_0\Delta_i M I_{\{(\Delta_i X)^2\le r_h,(\Delta_i L)^2\le 4r_h\}}}{\sqrt{h}}.\tag{42}$$

However, since both $P\{\frac{1}{\sqrt{h}}\sum_{i=1}^{n}\Delta_i X_0\Delta_i M I_{\{(\Delta_i X)^2\le r_h,(\Delta_i L)^2\le 4r_h,\Delta_i N\ne 0\}}\ne 0\}$ and $P\{\frac{1}{\sqrt{h}}\times\sum_{i=1}^{n}\Delta_i X_0\Delta_i M I_{\{(\Delta_i X)^2\le r_h,(\Delta_i M)^2\le 4r_h,\Delta_i N\ne 0\}}\ne 0\}$ are dominated by $nP\{\Delta_i N\ne 0,(\Delta_i M)^2>cr_h\}\to 0$, we have

$$\begin{aligned}
&\lim_h\frac{1}{\sqrt{h}}\sum_{i=1}^{n}\Delta_i X_0\Delta_i M I_{\{(\Delta_i X)^2\le r_h,(\Delta_i L)^2\le 4r_h\}}\\
&\quad=\lim_h\frac{1}{\sqrt{h}}\sum_{i=1}^{n}\Delta_i X_0\Delta_i M I_{\{(\Delta_i X)^2\le r_h,(\Delta_i L)^2\le 4r_h,\Delta_i N=0\}}\\
&\quad=\lim_h\frac{1}{\sqrt{h}}\sum_{i=1}^{n}\Delta_i X_0\Delta_i M I_{\{(\Delta_i X)^2\le r_h,(\Delta_i M)^2\le 4r_h,\Delta_i N=0\}}\\
&\quad=\lim_h\frac{1}{\sqrt{h}}\sum_{i=1}^{n}\Delta_i X_0\Delta_i M I_{\{(\Delta_i X)^2\le r_h,(\Delta_i M)^2\le 4r_h\}}.
\end{aligned}$$

Moreover, by the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&\frac{\sum_{i=1}^n \int_{t_{i-1}}^{t_i} a_u \,\mathrm{d}u\, \Delta_i M I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 \le 4r_h\}}}{\sqrt{h}} \\
&\quad \le \frac{\sqrt{\sum_{i=1}^n (\int_{t_{i-1}}^{t_i} a_u \,\mathrm{d}u)^2}}{\sqrt{h}} \sqrt{\sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i M)^2 \le 4r_h\}}} \qquad (43)\\
&\quad \le c \sqrt{\sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i M)^2 \le 4r_h\}}},
\end{aligned}$$

which tends to zero in probability since, by Remark 3.2, as $h \to 0$,

$$E\left[\sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i M)^2 \le 4r_h\}}\right] = \int_0^T \int_{|x| \le 2r_h^{1/4}} x^2 \nu(\mathrm{d}x) = T\eta^2(r_h^{1/4}) \to 0. \qquad (44)$$

On the other hand,

$$\begin{aligned}
&\frac{1}{\sqrt{h}} \sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} \sigma_u \,\mathrm{d}W_u\right) \Delta_i M I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 \le 4r_h\}} \\
&\quad = \frac{1}{\sqrt{h}} \sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} \sigma_u \,\mathrm{d}W_u\right) \Delta_i M^{(h)} I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 \le 4r_h\}} \qquad (45)\\
&\qquad - \frac{1}{\sqrt{h}} \sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} \sigma_u \,\mathrm{d}W_u\right) h\, \mathrm{d}\big(2\sqrt[4]{r_h}\big) I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 \le 4r_h\}},
\end{aligned}$$

where, using the fact that $\int_{t_{i-1}}^{t_i} \sigma_u \,\mathrm{d}W_u$ and $\Delta_i M^{(h)}$ are martingale increments with zero quadratic covariation, the $L^1(\Omega)$-norm of the first right-hand term is bounded by

$$\sqrt{E\left[\frac{\sum_{i=1}^n (\int_{t_{i-1}}^{t_i} \sigma_u \,\mathrm{d}W_u)^2 (\Delta_i M^{(h)})^2}{h}\right]},$$

which is dealt with similarly as in (31) and tends to zero. Moreover,

$$\begin{aligned}
&E\left[\frac{1}{\sqrt{h}} \sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} \sigma_u \,\mathrm{d}W_u\right) h\, \mathrm{d}\big(2\sqrt[4]{r_h}\big) I_{\{(\Delta_i X)^2 \le r_h, (\Delta_i M)^2 \le 4r_h\}}\right] \\
&\quad = c\sqrt{h}\left[I_{\alpha \ne 1}\big(c + r_h^{(1-\alpha)/4}\big) + I_{\alpha = 1} \ln \frac{1}{r_h^{1/4}}\right]
\end{aligned}$$

$$\times E\left[\sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}\sigma_u\,\mathrm{d}W_u\right)I_{\{(\Delta_i X)^2\le r_h,(\Delta_i M)^2\le 4r_h\}}\right]$$

$$\le c\sqrt{h}\left[I_{\alpha\neq 1}\big(c+r_h^{(1-\alpha)/4}\big)+I_{\alpha=1}\ln\frac{1}{r_h^{1/4}}\right]\sqrt{E\left[\sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}\sigma_u\,\mathrm{d}W_u\right)^2\right]}\to 0.$$

Using the fact that

$$\frac{\sqrt{2hIQ}}{\sqrt{2/3\sum_i(\Delta_i X)^4 I_{\{(\Delta_i X)^2\le r_h\}}}}$$

tends to 1 in probability, treating $I_4(h)$ as in (42) and putting together the simplified version of $I_2(h)$, we obtain that $(\hat{IV}_h-IV)/\sqrt{2h\hat{IQ}_h}$ is the sum of a term which converges in distribution to an $N(0,1)$ r.v. plus a negligible term and a remainder

$$-\frac{\sum_{i=1}^{n}(\int_{t_{i-1}}^{t_i}\sigma_u\,\mathrm{d}W_u)^2 I_{\{(\Delta_i X)^2>r_h,(\Delta_i X_1)^2\le 4r_h\}}}{\sqrt{2hIQ}}+\frac{\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i X)^2\le r_h,(\Delta_i M)^2\le 4r_h\}}}{\sqrt{2hIQ}}. \quad (46)$$

(a) If $\alpha<1$, the first term of (46) is negligible with respect to $\frac{r_h^{1-\alpha/2}}{\sqrt{2hIQ}}$, in fact,

$$\frac{\sum_{i=1}^{n}(\int_{t_{i-1}}^{t_i}\sigma_u\,\mathrm{d}W_u)^2 I_{\{(\Delta_i X)^2>r_h,(\Delta_i X_1)^2\le 4r_h\}}}{r_h^{1-\alpha/2}}\le\frac{\sum_{i=1}^{n}h\ln(1/h)I_{\{(\Delta_i X)^2>r_h\}}}{r_h^{1-\alpha/2}},$$

where

$$E\left[\frac{\sum_{i=1}^{n}h\ln(1/h)I_{\{(\Delta_i X)^2>r_h\}}}{r_h^{1-\alpha/2}}\right]\le h^{1-\beta}\ln\frac{1}{h}\to 0.$$

Therefore, (46) can be written as

$$\frac{r_h^{1-\alpha/2}}{\sqrt{2hIQ}}\left[\mathrm{o}_P(1)+\frac{\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i X)^2\le r_h,(\Delta_i M)^2\le 4r_h\}}}{r_h^{1-\alpha/2}}\right]. \quad (47)$$

Using (37), Lemma 3.1(i) and Theorem 3.4, we arrive at

$$\frac{\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i X)^2\le r_h,(\Delta_i M)^2\le 4r_h\}}}{r_h^{1-\alpha/2}}$$

$$\le\frac{\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 9r_h/4\}}+\mathrm{o}_P(h^{1-\alpha/2})}{r_h^{1-\alpha/2}}$$

$$\sim\frac{\sum_{i=1}^{n}(\Delta_i M)^2 I_{\{(\Delta_i M)^2\le 9r_h/4\}}}{r_h^{1-\alpha/2}}$$

$$\le \frac{\sum_i (\int_{t_{i-1}}^{t_i} \int_{|x|\le 3\sqrt{r_h}/2} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t) - h\int_{3\sqrt{r_h}/2<|x|\le 1} x\nu(\mathrm{d}x))^2}{r_h^{1-\alpha/2}}$$

$$= R_h + Tc + Tc\left(\frac{h}{r\,h}\right)^{\alpha/2} h^{1-\alpha/2} \xrightarrow{P} Tc,$$

where the term R_h has variance $\sim c r_h^{\alpha/2} \to 0$ and so converges to zero in probability. Since $\frac{r_h^{1-\alpha/2}}{\sqrt{h}} \to 0$, we arrive at

$$\frac{\hat{IV}_h - IV}{\sqrt{2h\hat{IQ}_h}} \xrightarrow{st} N(0,1).$$

(b) If $\alpha > 1$, define $R_t := \sum_{s\le t} I_{\{|\Delta M_s|>\sqrt{h}\}}$. Then, by (37), the last term (times $\sqrt{2IQ}$) in (46) dominates

$$\begin{aligned}
&\frac{\sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i M)^2\le r_h/16\}} - \mathrm{o}_P(h^{1-\alpha/2})}{\sqrt{h}}\\
&\quad= \frac{1}{\sqrt{h}}\left[\sum_i (\Delta_i M)^2 I_{\{\Delta_i R=0\}} + \sum_i (\Delta_i M)^2 \left[I_{\{(\Delta_i M)^2\le r_h/16\}} - I_{\{\Delta_i R=0\}}\right]\right] - \mathrm{o}_P(h^{1/2-\alpha/2})\\
&\quad\ge -\mathrm{o}_P(h^{1/2-\alpha/2}) + \frac{\sum_i (\int_{t_{i-1}}^{t_i}\int_{|x|\le\sqrt{h}} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t) - h\int_{\sqrt{h}<|x|\le 1} x\nu(\mathrm{d}x))^2}{\sqrt{h}}\\
&\qquad - \frac{\sum_i (\Delta_i M)^2 I_{\{(\Delta_i M)^2 > r_h/16, \Delta_i R=0\}}}{\sqrt{h}}.
\end{aligned}\tag{48}$$

First,

$$\begin{aligned}
&\sum_i (\Delta_i M)^2 I_{\{(\Delta_i M)^2>r_h/16,\Delta_i R=0\}}\\
&\quad= \sum_i \left[\Delta_i[M] + 2\int_{t_{i-1}}^{t_i} (M_{s_-} - M_{t_{i-1}})\,\mathrm{d}M_s\right] I_{\{(\Delta_i M)^2>r_h/16,\Delta_i R=0\}}.
\end{aligned}$$

As in Lemma 3.1, the sum of the right-hand terms within brackets is of order $u_n = (n/\log n)^{1/\alpha}$ so that

$$\frac{\sum_i |\int_{t_{i-1}}^{t_i} (M_{s_-} - M_{t_{i-1}})\,\mathrm{d}M_s|}{\sqrt{h}} = \frac{u_n \sum_i |\int_{t_{i-1}}^{t_i} (M_{s_-} - M_{t_{i-1}})\,\mathrm{d}M_s|}{u_n\sqrt{h}} \xrightarrow{P} 0$$

since $u_n\sqrt{h} = (\frac{n^{(1-\alpha/2)}}{\log n})^{1/\alpha} \to +\infty$. Theorem 3.4 applied with $u = 1/2$ yields that with $\varepsilon = h^{1/2}$,

$$\sum_i \left(\int_{t_{i-1}}^{t_i}\int_{|x|\le\sqrt{h}} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t) - h\int_{\sqrt{h}<|x|\le 1} x\nu(\mathrm{d}x)\right)^2 = \varepsilon^{2-\alpha/2} Y_h + Tc\varepsilon^{2-\alpha} + Tch\varepsilon^{2-2\alpha},$$

where $\mathrm{var}(Y_h) \to 1$. Therefore, in (48), we remain with

$$h^{1/2-\alpha/2}\left[-\mathrm{o}_P(1) + h^{\alpha/4}Y + Tc + Tch^{1-\alpha/2} - \frac{\sum_i \Delta_i[M]I_{\{(\Delta_i M)^2 > r_h/16, \Delta_i R=0\}}}{h^{1-\alpha/2}}\right] \stackrel{a.s.}{\to} +\infty,$$

where the divergence is due to the fact that $h^{1/2-\alpha/2} \to +\infty$ while $\frac{\sum_i \Delta_i[M]I_{\{(\Delta_i M)^2 > r_h/16, \Delta_i R=0\}}}{h^{1-\alpha/2}}$ tends to zero in probability since its expected value is dominated by

$$\begin{aligned}&\frac{n}{h^{1-\alpha/2}} E^{1/2}\big[\big(\Delta_i[M]I_{\{\Delta_i R=0\}}\big)^2\big] P^{1/2}\{(\Delta_i M)^2 > r_h/16, \Delta_i R = 0\}\\ &\quad\le \frac{n}{h^{1-\alpha/2}}\left(h\int_{|x|\le\sqrt{h}} x^4\nu(\mathrm{d}x)\right)^{1/2} h^{(2-\alpha/2-\beta)1/2} = h^{(1-\beta)/2} \to 0,\end{aligned}$$

having used the fact that

$$\begin{aligned}P\{(\Delta_i M)^2 > r_h, \Delta_i R = 0\} &= P\big\{(\Delta_i M)^2 I_{\{\Delta_i R=0\}} > r_h\big\} \le \frac{E[(\Delta_i M)^2 I_{\{\Delta_i R=0\}}]}{r_h}\\ &= \frac{h\int_{|x|\le\sqrt{h}} x^2\nu(\mathrm{d}x)}{r_h} = h^{2-\alpha/2-\beta}.\end{aligned} \tag{49}$$

On the other hand, the first term in (46) is negligible with respect to $h^{1/2-\alpha/2}$ (the speed of divergence of $(\sum_{i=1}^n (\Delta_i M)^2 I_{\{(\Delta_i M)^2 \le r_h/16\}} - \mathrm{o}_P(h^{1-\alpha/2}))/\sqrt{h})$ because

$$\frac{\sum_{i=1}^n (\int_{t_{i-1}}^{t_i} \sigma_u\,\mathrm{d}W_u)^2 I_{\{(\Delta_i X)^2 > r_h, (\Delta_i X_1)^2 \le 4r_h\}}}{\sqrt{h}h^{1/2-\alpha/2}} \le \frac{h\log(1/h)h^{-\alpha\beta/2}}{h^{1-\alpha/2}} = h^{(\alpha/2)(1-\beta)}\log\frac{1}{h} \to 0.$$

Therefore, (46) explodes to $+\infty$. Finally, if $\alpha = 1$ in (46), then the first term is negligible, as

$$\frac{\sum_{i=1}^n (\int_{t_{i-1}}^{t_i} \sigma_u\,\mathrm{d}W_u)^2 I_{\{(\Delta_i X)^2 > r_h, (\Delta_i X_1)^2 \le 4r_h\}}}{\sqrt{h}} = \mathrm{O}_p\left(h^{(1-\beta)/2}\log\frac{1}{h}\right) \to 0.$$

For the second term, we take a $\delta > 0$ such that $2/3 < \beta + \delta < 1$, we choose $\varepsilon = h^{(\beta+\delta)/2}$ and we use the same steps as were used to reach (48) for $\alpha > 1$, but we consider $\tilde{R}_t = \sum_{s\le t} I_{\{|\Delta M_s| > \varepsilon\}}$ in place of R_t. Also using Theorem 3.4, we obtain that the second term in (46) dominates

$$\begin{aligned}&\frac{Y_h\varepsilon^{3/2}}{\sqrt{2hIQ}} + \frac{\varepsilon}{\sqrt{2hIQ}} - \frac{\sum \Delta_i[M]I_{\{(\Delta_i M)^2 > r_h/16, \Delta_i \tilde{R}=0\}}}{\sqrt{2hIQ}}\\ &\quad - \frac{2\sum_{i=1}^n \int_{t_{i-1}}^{t_i} (M_{s-} - M_{t_{i-1}})\,\mathrm{d}M_s\, I_{\{(\Delta_i M)^2 > r_h/16, \Delta_i \tilde{R}=0\}}}{\sqrt{2hIQ}},\end{aligned}$$

where the variance of Y_h tends to 1 so that $Y_h\varepsilon^{3/2}/\sqrt{h}$ tends to zero in probability. The second term tends to $+\infty$ at rate $\varepsilon/\sqrt{h}$. The third term is negligible with respect to $\varepsilon/\sqrt{h}$: applying (49) with $\tilde{R}$ in place of R and the Cauchy–Schwarz inequality, we get

$$E\left[\frac{1}{\varepsilon}\sum\int_{t_{i-1}}^{t_i}\int_{|x|\le 1} x^2\mu(\mathrm{d}x,\mathrm{d}t) I_{\{(\Delta_i M)^2>r_h/16,\Delta_i\tilde{R}=0\}}\right] = \mathrm{O}(h^{\delta/2}) \to 0.$$

Finally, the last term is also negligible since the speed of convergence to zero of the numerator is $u_n = n/\log^2 n$ (as in the proof of Lemma 3.1) and $u_n\sqrt{h} \to +\infty$. So, even for $\alpha = 1$, the normalized bias $(\hat{IV}_h - IV)/\sqrt{2h\hat{IQ}_h}$ diverges to $+\infty$. □

Proof of Proposition 4.1. As in Lemma A.2 with $\sqrt{r_h}$ in place of r_h as bound for $\max_{i=1,\dots,n}|a_{ni}|$, using the fact that $\alpha < 1$ and applying Lemma 3.1(i), we deduce that $\hat{H}_h$ has the same limit in probability as

$$X_T - \sum_{i=1}^n (\Delta_i X_0 + \Delta_i M) I_{\{\Delta_i N=0,(\Delta_i M)^2\le r_h\}}$$

when $h \to 0$. Moreover, since a.s. $N_T < \infty$ and $\sum_{i=1}^n \Delta_i X_0 I_{\{(\Delta_i M)^2>r_h\}}) = \mathrm{O}_P(h^{(1-\alpha\beta)/2} \times \sqrt{\log(1/h)}) \to 0$, taking $\tilde{R}_t = \sum_{s\le t} I_{\{|\Delta M_s|>\sqrt{r_h}\}}$, the above term has limit in probability equal to

$$\begin{aligned}
&X_T - \lim_h \sum_{i=1}^n \bigl(\Delta_i X_0 + \Delta_i M I_{\{(\Delta_i M)^2\le r_h\}}\bigr) \\
&\quad = X_T - X_{0T} - \lim_h \left[\sum_{i=1}^n \int_{t_{i-1}}^{t_i}\int_{|x|\le\sqrt{r_h}} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t) - T\int_{\sqrt{r_h}<|x|\le 1} x\nu(\mathrm{d}x)\right] \\
&\qquad - \lim_h \sum_i \Delta_i M\bigl(I_{\{(\Delta_i M)^2\le r_h\}} - I_{\{\Delta_i\tilde{R}=0\}}\bigr).
\end{aligned}$$

Using the fact that $P\{\Delta_i\tilde{R} \ge 1\} = \mathrm{O}(h^{1-\alpha\beta/2})$, as was used after (25), we deduce that $\sum_i \Delta_i M I_{\{(\Delta_i M)^2\le r_h,\Delta_i\tilde{R}\ge 1\}} = \mathrm{O}_P(h^{(1-\alpha)\beta/2}) \to 0$. Using the Hölder inequality with exponents $p = q = 2$, we have $\sum_i \Delta_i M I_{\{(\Delta_i M)^2>r_h,\Delta_i\tilde{R}=0\}} = \mathrm{O}_P(r_h^{(1-\alpha)\beta/2}) \to 0$. Finally, $\int_0^T\int_{|x|\le\sqrt{r_h}} x\tilde{\mu}(\mathrm{d}x,\mathrm{d}t) \xrightarrow{L^2} 0$ and $\int_{\sqrt{r_h}<|x|\le 1} x\nu(\mathrm{d}x) \to m$ so that $\hat{H}_{h,T} \xrightarrow{P} J_T + mT$. □

Acknowledgements

A previous version of this working paper appeared as "Nonparametric test for analyzing the fine structure of price fluctuations", Columbia Financial Engineering Report 2007-13. This research was supported in part by the European Science Foundation program "Advanced Mathematical Methods in Finance", by Istituto Nazionale di Alta Matematica and by MIUR Grants Nos 206132713-001 and 2004011204-002. We thank Jean Jacod and Suzanne Lee for important comments.

References

[1] Aït-Sahalia, Y. (2004). Disentangling volatility from jumps. *J. Financial Economics* **74** 487–528.
[2] Aït-Sahalia, Y. and Jacod, J. (2009). Testing for jumps in a discretely observed process. *Ann. Statist.* **37** 184–222. MR2488349
[3] Aït-Sahalia, Y. and Jacod, J. (2009). Estimating the degree of activity of jumps in high frequency data. *Ann. Statist.* **37** 2202–2244. MR2543690
[4] Aït-Sahalia, Y. and Jacod, J. (2009). Testing whether jumps have finite or infinite activity. Working paper.
[5] Alili, L. and Kyprianou, A. (2005). Some remarks on first passage of Lévy processes, the American put and pasting principles. *Ann. Appl. Probab.* **15** 2062–2080. MR2152253
[6] Barndorff-Nielsen, O.E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *J. Financial Econometrics* **4** 1–30. MR2222740
[7] Barndorff-Nielsen, O.E., Shephard, N. and Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Process. Appl.* **116** 796–806. MR2218336
[8] Carr, P., Geman, H., Madan, D. and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *J. Business* **75** 305–332.
[9] Carr, P. and Wu, L.R. (2003). What type of process underlies options? A simple robust test. *J. Finance* **LVIII** 2581–2610.
[10] Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Boca Raton, FL: CRC Press. MR2042661
[11] Ikeda, N. and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. Amsterdam: North Holland. MR1011252
[12] Jacod, J. (2004). The Euler scheme for Lévy driven stochastic differential equations: Limit theorems. *Ann. Probab.* **32** 1830–1872. MR2073179
[13] Jacod, J. (2008). Asymptotic properties of realized power variations and associated functions of semimartingales. *Stochastic Process. Appl.* **118** 517–559. MR2394762
[14] Jacod, J. (2007). Statistics and high-frequency data. In *SEMSTAT 2007*, La Manga, Spain, May 2007.
[15] Jacod, J. and Protter, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307. MR1617049
[16] Kou, S. (2002). A jump-diffusion model for option pricing. *Management Science* **48** 1086–1101.
[17] Lee, S. and Mykland, P.A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Rev. Financial Stud.* **21** 2535–2563.
[18] Madan, D.B. (2001). Purely discontinuous asset price processes. In *Option Pricing, Interest Rates and Risk Management* (J. Cvitanic, E. Jouini and M. Musiela, eds.) 105–153. Cambridge: Cambridge Univ. Press. MR1848551
[19] Mancini, C. (2004). Estimation of the parameters of jump of a general Poisson-diffusion model. *Scand. Actuar. J.* **1** 42–52. MR2045358
[20] Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Statist.* **36** 270–296. MR2528985
[21] Mancini, C. and Renó, R. (2008). Threshold estimation of Markov models with jumps and interest rate modeling. *J. Econometrics*. To appear. DOI:10.1016/j.jeconom.2010.03.019.
[22] Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *J. Financial Economics* **3** 125–144.
[23] Metivier, M. (1982). *Semimartingales: A Course On Stochastic Processes*. Berlin: de Gruyter. MR0688144
[24] Protter, P.E. (2005). *Stochastic Integration and Differential Equations*. Berlin: Springer. MR2273672

[25] Todorov, V. and Tauchen, G. (2010). Activity signature functions for high-frequency data analysis. *J. Econometrics* **154** 125–138. MR2558956
[26] Woerner, J. (2006). Analyzing the fine structure of continuous time stochastic processes. Univ. Göttingen. Working paper.

Received May 2009 and revised March 2010