What you don't know cannot hurt you: On the detection of small jumps.

Lan Zhang

This version: September 16, 2007

Abstract

With the availability of high frequency financial data, nonparametric estimation of volatility of an asset return process becomes feasible. One of the main issues is concerned with separating the volatility due to continuous evolution from the volatility due to jumps. This paper argues that the separation of continuous and jump evolution is soft, in addition, large biases can occur when estimating volatility in the presence of infinitely many small jumps. The paper advocates that for the purposes of application and analysis, one can identify small jumps with continuous evolution while treating large jumps as compound Poisson type. A main contiguity theorem is derived to relate compound poisson process to Lévy process with infinite activity in the asymptotic setting, and gives the cutoff below which the jumps are too small to be detectable. The theorem provides a straightforward tool to simulate a Lévy process with infinite activity, it also reveals why a "good" estimator can behave badly in finite samples.

Some key words and phrases: Blumenthal-Getoor index, compound Poisson process, consistency, contiguity, discrete observation, infinite activity process, jumps, Lévy process, rate of convergence, volatility.

1 Introduction

With the availability of high frequency financial data, nonparametric estimation of volatility of an asset return process becomes feasible. One of the main issues which has emerged in the literature is concerned with separating the volatility due to continuous evolution from the volatility due to jumps. (Aït-Sahalia (2002), Aït-Sahalia and Jacod (2004, 2007)), Barndorff-Nielsen and Shephard (2004, 2005, 2006), Barndorff-Nielsen, Shephard, and Winkel (2006), Fan and Wang (2005), Huang and Tauchen (2006), Lee and Mykland (2006), and Mancini (2004, 2006)).

From this literature, it has become clear that consistency is usually available under broad conditions. However, it is hard to obtain sharp results for estimators of continuous volatility. Due to the lack of sharp convergence rates and central limit theorems in the literature, confidence intervals for estimators are frequently unavailable.

This raises the question of whether such theory development is unavailable because of analytical difficulty, or because of more fundamental problems on the part of the estimators.

The current paper argues that these two issues are intertwined. On the one hand, the separation between continuous and jump evolution is soft: while large jumps are easily separable from other evolution, small jumps are both theoretically and practically difficult to separate from continuous evolution. In particular, one shall see that large biases can occur when estimating volatility in the presence of infinitely many small jumps.

On the other hand, the paper advocates that one can easily live with it by simply identifying small jumps with the continuous part of the process. Application such as trading, for example, is unaffected by the separation problem. The upside is even greater for the purposes of econometric analysis. By identifying small jumps with continuous evolution, general jump processes can be treated as if they were of the compound Poisson type. This provides a new tool which can be used both for Monte Carlo simulation, and to show theoretical properties of the estimators.

To illustrate the usefulness of the approach, this paper provides a central limit theory for a class of multi-power estimators for α-stable processes with $1 < \alpha < 3/2$. To my best knowledge, this is the first limit theory in existence for estimators of continuous volatility when α is greater than one, and unknown.

The aim of this paper is mainly conceptual, and we have therefore not issues such as microstructure. It is conjectured that microstructure will make the separation between continuous and jumps evolution even more difficult.

The paper is organized as follows. Section 2 introduces the problem of estimating the continuous part of the volatility of a process. Section 3 develops the main contiguity theorem and discusses the threshold below which jumps are unidentifiable. This gives rise to a new tool for both simulation and theoretical computation identified, as seen in Section 4. Section 5 derives the asymptotic distribution of multi-power estimators. These estimators are (under some conditions) asymptotically unbiased at the desirable $\sqrt{n}$ rate. Finally Section 6 reveals how the bias induced by the separation difficulty can be characterized in an asymptotic description.

2 Estimation of continuous volatility

Consider a process $\{X_t\}$, such as stock price series at logarithmic scale, on the form

$$X_t = X_0 + \int_0^t \mu_u du + \int_0^t \sigma_u dW_u + J_t, \tag{1}$$

where W is a Brownian motion and J is a process which only evolves through jumps. The continuous part of X, $\int_0^t \mu_u du + \int_0^t \sigma_u dW_u$, has drift μ_u and volatility σ_u.

The typical estimation problem is as follows. The object of inference is the quadratic variation of the continuous part of X, defined by

$$[X^c, X^c]_T = \int_0^T \sigma_u^2 du. \tag{2}$$

Inference is based on observations that take place at the grid of time points $\mathcal{G}_n = \{t_{n,i}, i = 0, 1, 2, \cdots n\}$ that span the time interval $[0, T]$. When analyzing the asymptotics of an estimator for (2), one lets $\mathcal{G}_n$ become dense in $[0, T]$ as $n \to \infty$, while T remains fixed.

It is known in the literature that $[X^c, X^c]_T$ can be consistently estimated (cf. the papers cited in the Introduction). Moreover, when $\{J_t\}$ has finitely many jumps, estimators $\widehat{[X^c, X^c]}_T$ are typically $\sqrt{n}$-consistent, that is, $\widehat{[X^c, X^c]}_T - [X^c, X^c]_T = O_p(n^{-1/2})$. Note that for a pure jump Lévy process $\{J_t\}$, having finitely many jumps is equivalent to being compound Poisson.

The problem arises when there are infinitely many jumps. For example, consider

$$J_t = \text{ symmetric or asymmetric stable process with index } \alpha, \tag{3}$$

with $0 < \alpha < 2$. In other words, J_t is Lévy process with no continuous component, with Lévy measure

$$\nu(dx) = \begin{cases} c^+ x^{-\alpha-1} dx, & \text{if } x > 0 \\ c^- |x|^{-\alpha-1} dx, & \text{if } x < 0 \end{cases} \tag{4}$$

cf. p. 217 of Bertoin (1998). See also Chapter 1 of Protter (2004) on Lévy processes. The parameter α is unknown, and is usually called the Blumenthal-Getoor index.

It seems that when $\alpha < 1$, inference remains well behaved in the sense that estimators are $\sqrt{n}$-consistent and analyses can be carried through fairly easily. Both Aït-Sahalia and Jacod (2007) and Barndorff-Nielsen, Shephard, and Winkel (2006) show $\sqrt{n}$ consistency for their respective estimators for $[X^c, X^c]_T$.

However, in the case when $1 < \alpha < 2$, there are almost no results available on limit laws, nor on sharp rates of convergence. To my best knowledge, the literature displays no $\sqrt{n}$-consistent estimator for $[X^c, X^c]_T$ when J_t is a stable process with unknown index α. This is even in the case when σ_t is taken to be constant, let alone a more general Lévy process.

This raises the possibility that when $1 < \alpha < 2$, the rate of convergence for estimating $[X^c, X^c]_T$ can be slower than that for the total quadratic variation of the process X_t. To see this, note that

$$\begin{aligned}\underbrace{[X,X]_T}_{\text{total quadratic variation}} &= \underbrace{\int_0^T \sigma_u^2 du}_{[X^c,X^c]_T} + \sum_{0\le s\le T} (\Delta J_s)^2 \\ &= \lim_{n\to\infty} \underbrace{\sum_{i=1}^{n} (X_{t_{n,i}} - X_{t_{n,i-1}})^2}_{\text{realized volatility}}. \end{aligned} \tag{5}$$

The total quadratic variation can be estimated by the realized volatility in a $\sqrt{n}$-consistent manner, see Section 6 of Jacod and Protter (1998), and more recently Jacod (2006).

Things are not as bad as they might have been. This paper shows that there exist $\sqrt{n}$-consistent estimators for $\alpha > 1$, and we provide an instance of this in section 5. However, the $\sqrt{n}$ rate fails to hold uniformly. This latter problem is discussed in Section 6.

REMARK 1. In the case where $\{J_t\}$ follows compound Poisson (finitely many jumps), estimators for $[X^c, X^c]_T$ remain $\sqrt{n}$-consistent. This can be seen from the approach of Mancini (2004, 2006) (see also Fan and Wang (2005) and Lee and Mykland (2006)). With this technique, those finitely many intervals $[t_{n,i-1}, t_{n,i}]$ that contain jumps can be located and removed before further inference, without affecting any convergence rates. In some sense, the fact that there are finitely many jumps means that all jumps can be detected for sufficiently large n (thus sufficiently finer intervals). The technique can be used to define estimators for $[X^c, X^c]_T$, or to analyze other existing estimators, such as those of Aït-Sahalia and Jacod (2004, 2007)), Barndorff-Nielsen and Shephard (2004, 2005, 2006), and Huang and Tauchen (2006).

Specifically when $n \to \infty$, $X_{t_{n,i}} - X_{t_{n,i-1}}$ is either approximately $\sigma_{t_{n,i-1}}$N(0,1) (when there is no jump in $(t_{n,i}, t_{n,i-1}])$, or approximately $\sigma_{t_{n,i-1}}$N(0,1) + jump (when there is a jump in the interval). As n grows large, the magnitude of the latter is increasingly big relative to the former, and can be detected at an exponential rate.

Note that in the case where J_t is an α-stable process with known α, it is documented in Aït-Sahalia and Jacod (2007) that $\sqrt{n}$-consistent estimation is possible. Our arguments concern the situation where J_t has an unknown distribution (unknown α). □

3 Indistinguishable jumps

In this paper, we consider X process satisfying (1), where μ and σ are constant, where the pure jump component has Lévy measure ν. The assumption on constant (μ, σ) are not necessary but will nicely simplify the presentation. Moreover, we assume that X is observed at nonrandom times which are evenly spaced across $[0, T]$.

Definition 1. *We consider two classes of Lévy measures.* Compound Poisson measures *are ν's that give zero measure to an open interval containing zero.* Regular infinite activity measures *are ν's that are continuously differentiable around (but not at) zero, with*

$$\frac{\nu'(x)}{|x|^{-\alpha-1}} \rightarrow \begin{cases} c^{+} & as\ x \downarrow 0 \\ c^{-} & as\ x \uparrow 0 \end{cases} \tag{6}$$

for some Blumenthal-Getoor index $\alpha \in (0, 2)$, where c^{+} and c^{-} are both positive. We suppose that for the same α, $\sup_{x \neq 0} \nu'(x)/|x|^{-\alpha-1} < \infty$.

To see the difficulty in distinguishing small jumps from continuous evolution of X, we construct a division of J_t into small jumps and large jumps processes.

Definition 2. *For $h > 0$, let*

$$G(h) = \sup\{h' \geq 0 : \ E \sum_{0<u\leq 1} \Delta J_u I_{\{h > \Delta J_u > -h'\}} \geq 0\}.$$

Also set

$$k = k_\alpha = \left(\frac{c^{+}}{c^{-}}\right)^{1/(1-\alpha)}. \tag{7}$$

Let h be a nonrandom cutoff ($h > 0$). Set

$$\begin{aligned} J_t^{(l)} &= \sum_{0<u\leq t} \Delta J_u I_{\{\Delta J_u > h \text{ or } \Delta J_u < -G(h)\}} \\ J_t^{(s)} &= J_t - J_t^{(l)}. \end{aligned} \tag{8}$$

Thus, small jumps are those that jump within the band of $[-G(h), h]$. Basic facts about this split are as follows:

Proposition 1. *Let ν be a regular infinite activity Lévy measure with Blumenthal-Getoor index $\alpha \in (0, 2)$.*

(i) $\{J_t^{(l)}\}$ and $\{J_t^{(s)}\}$ are Lévy processes. $\{J_t^{(s)}\}$ has moments of all orders. Also $\{J_t^{(s)}\}$ is a martingale , and

$$Var(J_T^{(s)}) = \langle J^{(s)}, J^{(s)} \rangle_T = Tc^{+}(1 + k_\alpha)(2-\alpha)^{-1} h^{2-\alpha}(1 + o(1)) \tag{9}$$

as $h \to 0$.

(ii) The function $G(h)$ is nondecreasing, and increasing when $\nu'(h) > 0$. $G(h)/h \to k_\alpha$ as $h \downarrow 0$. In the α stable case (ν satisfies (3)-(4)), $G(h) = k_\alpha h$ for all $h > 0$. The equation (10) is in this case exact for all h (the $o(1)$ term is zero).

PROOF OF PROPOSITION 1: By Theorems 36-37 (p. 26-27) in Protter (2004), $J^{(l)}$ and $J^{(s)}$ are Lévy processes. Also, the small-jump process $J^{(s)}$ has moments of all orders, and is a martingale, following Theorem 34 (p. 25) and similar derivation of Theorem 41 (p. 30), respectively, in Protter (2004)). Finally,

$$\begin{aligned}\mathrm{Var}(J_T^{(s)}) &= \langle J^{(s)}, J^{(s)} \rangle_T \\ &= T(c^+ + c^- k_\alpha^{2-\alpha})(2-\alpha)^{-1} h^{2-\alpha}(1+o(1)) \\ &= Tc^+(1+k_\alpha)(2-\alpha)^{-1} h^{2-\alpha}(1+o(1)),\end{aligned} \tag{10}$$

where the second line follows from the same methods as in (A.5)-(A.6) in the Appendix. This shows *(i)*. The martingaleness of $J^{(s)}$ also means that $E\sum_{0<u\leq 1} \Delta J_u I_{\{h>\Delta J_u > -G(h)\}} = 0$, whence *(ii)* follows. ∎

Two probability distributions of Levy processes are given by

$$\begin{aligned} P: &\quad X_t = X_0 + \mu t + \sigma W_t + J_t \\ P_h: &\quad X_t = X_0 + \mu_h t + \sigma_h W_t + J_t^{(l)}, \end{aligned} \tag{11}$$

where $\sigma_h^2 = \sigma^2 + \mathrm{Var}(J_1^{(s)})$. (11) says that under P, the jump component J follows a Lévy process which contains both large jumps and infinitely many small jumps, whereas under P_h, the jump component $J^{(l)}$ follows a compound Poisson which contains only large jumps while (infinitely many) small changes in X are treated as part of the continuous evolution in X. One can view P as a true model governing the data-generating process and P_h as a working model due to its easy analysis, or one can consider P and P_h as two competing models for X.

Note that the jump components in both P and P_h are Lévy processes, with Blumenthal-Getoor index $0 < \alpha < 2$ (unknown) and $\alpha = 0$, respectively. The Lévy measure for $J^{(l)}$ under P_h is

$$\nu_h(dx) = \nu(dx) I(x > h, \text{ or } x < -k_\alpha h).$$

We now have the following theorem, with interpretation to follow:

Theorem 1. *Let ν be a regular infinite activity Lévy measure with Blumenthal-Getoor index $\alpha \in (0,2)$. Suppose that $\sigma^2 > 0$. Let the observation times be equidistant, $t_{n,i} = iT/n$. Consider P and P_h as probabilities on the observables $X_{t_{n,i}}$. Let $h = h_n = h_0 n^{-\gamma}$, where h_0 and γ are positive. Then, if $\gamma > \min(1/(3-\alpha), 1/2)$, we have*

$$\frac{dP_{h_n}}{dP}(X_0, X_{t_{n,1}}, ..., X_T) \to 1 \tag{12}$$

in P-probability as $n \to \infty$; if $\gamma = 1/(3-\alpha)$ and $\alpha > 1$, we have

$$\frac{dP_{h_n}}{dP}(X_0, X_{t_{n,1}}, ..., X_T) = O_{ui}(1) \tag{13}$$

in P-probability, as $n \to \infty$.

"$O_{ui}(1)$" means that the quantity is uniformly integrable. That is, P_{h_n} and P are contiguous, this is slightly stronger than $O_p(1)$. For more readings on contiguity, see Hájek and Sidak (1967) (Chapter VI), LeCam (1986), LeCam and Yang (1986), and Jacod and Shiryaev (2003) (Chapter IV) .

It is important to see that P and P_{h_n} agree on large jumps, but differ on (infinitely many) small jumps. Theorem 1 says that on the basis of finite amount of data, it is difficult to distinguish between models P and P_{h_n}. In particular, it is difficult to determine whether a Lévy process has infinite jump activity or not. In other words, small jumps are hardly separable from a continuous evolution in X, given fixed amount of data. Of course, as one gets more data, the cutoff h_n moves closer to zero, thus more jumps become "large" enough to be detected. It is the jumps below the cutoff $h_n = h_0 n^{-\gamma}$ that cannot be picked up. We can call these jumps, $J^{(s)}$, *undetectable.*

From (10) and (11), one can also see that

$$\begin{aligned}\sigma^2_{h_n} T - \sigma^2 T = \langle J^{(s)}, J^{(s)} \rangle_T &= Tc^+(1+k_\alpha)(2-\alpha)^{-1} h_n^{2-\alpha} \\ &= n^{-(2-\alpha)\gamma} Tc^+(1+k_\alpha)(2-\alpha)^{-1} h_0^{2-\alpha}. \end{aligned} \tag{14}$$

Hence, if $\alpha > 1$ and $\gamma = 1/(3-\alpha)$, $\langle J^{(s)}, J^{(s)} \rangle_T$ be bigger than $O(n^{-1/2})$, hence the undetectable jumps have a contribution to the total variance which is greater than the typical $O(n^{-1/2})$ random estimation error. This seems to suggest that separating small jumps from X has a statistical cost, namely, it incurs greater uncertainty. The matter is pursued further in Section 6.

REMARK 2. (Symmetric processes). It can be seen from the proof of Theorem 1 (and 5) that $O_{ui}(1)$ term in (13) reflects asymmetry in jumps. If J_t is a symmetric L'evy process (so $\nu'(x)$ is an even function),the $O_{ui}(1)$ terms in (13) and (A.3) are of the form $1 + o_p(1)$. For a symmetric process J, one can obtain (13) under the condition of $\alpha > 1$ and $\gamma = \frac{3}{2(4-\alpha)}$ (for details, see argument as in the proof of Theorem 5). The non-vanishing terms in equation (A.12) are the ones relating to the fourth cumulant, and its square. Hence, in the symmetric case, even greater jumps can be neglected. □

A natural question is whether a distinction (between small jumps and continuous evolution) that is so hard to make on the basis of data can actually have any importance in applications? A trading strategy that involves finite number of transactions cannot be affected by jumps of size less than $h = h_0 n^{-1}$. This seems to be a case where what you don't know cannot hurt you.

4 Application to computation and simulation

A direct application of Theorem 1 is to provide a straightforward simulation procedure for a Lévy process. This is achieved with the help of approximation model P_h.

If one wishes to simulate at times $t_{n,i} = iT/n$, under P, one can instead simulate under P_h. The jumps of the process under P_h are of compound Possion type. Let ν_h be the Lévy measure of X_t under P_h. Then

$$\nu_h(dx) = \begin{cases} v(dx) \text{ if } x > h \text{ or } x < -G(h) \\ 0 \text{ otherwise} \end{cases} \tag{15}$$

The Poisson intensity is given by

$$\lambda_h = \nu_h(\mathbb{R}). \tag{16}$$

The probability distribution Q of the jumps is given by

$$Q(dx) = \nu_h(dx)/\nu_h(\mathbb{R}). \tag{17}$$

This gives the general scheme for simulating under P_h.

More specifically when ν is α-stable,

$$\begin{aligned} \lambda_h = \nu_h(\mathbb{R}) &= (c^+ + c^- k_\alpha^{-\alpha}) \int_h^\infty x^{-\alpha-1} dx \\ &= \frac{c^+ + c^- k_\alpha^{-\alpha}}{\alpha} h^{-\alpha}. \end{aligned} \tag{18}$$

Further, $Q(\text{jump } > h) = c^+/(c^+ + c^- k_\alpha^{-\alpha})$. If F_h is the distribution functions of the positive jumps, one gets $F_h(x) = F_1(x/h)$, with $F_1(x) = \max(1 - x^{-\alpha}, 0)$. Obviously, after being suitably scaled, the negative jumps have the same law as minus the positive jumps. In other words, for $x > 0$, $Q(\text{jump } \leq -x | \text{ jump } < 0) = 1 - F_1(x/hk_\alpha)$. Finally, the inverse function $F_1^{(-1)}(y) = h(1-y)^{-1/\alpha}$.

The procedure for simulating the process X_t at times $t_{n,i} = iT/n$ under P_h is as follows:

(i) Simulate the continuous part $\mu t + \sigma_h W_t$ using normal random variables.

(ii) Simulate the jump times of X_t using exponential random variables with Poisson intensity λ_h. Let N_T be the number of jumps.

(iii) Simulate the size of the jumps as follows. Let U_i, $i = 1, ..., N_T$ be iid $(0, 1)$ uniform random variables. Let S_i be iid Bernoulli random variable, with value either 1 or -1 with probability $c^+/(c^+ + c^- k_\alpha^{-\alpha})$ and $c^- k_\alpha^{-\alpha}/(c^+ + c^- k_\alpha^{-\alpha})$, respectively. Jump size is then given as

$$\text{size of jump } \#i = \begin{cases} hU_i^{-1/\alpha}, & \text{if } S_i = 1 \\ -hk_\alpha U_i^{-1/\alpha}, & \text{if } S_i = -1 \end{cases} \tag{19}$$

To see (iii), note that $Z_{i,1} = U_i^{-1/\alpha} = F_1^{(-1)}(1 - U_i)$ has distribution F_1, while $Z_{i,h} = hU_i^{-1/\alpha}$ has distribution F_h. Thus (19) has distribution Q.

5 Application to bi- and multi-power estimators

From now on, the Lévy measure is assumed to be on the α-stable form. Consider the multipower estimator of $[X^c, X^c]_T = \sigma^2 T$ (Barndorff-Nielsen and Shephard (2004, 2005, 2006)). The q-th power estimator is given by

$$\widehat{[X^c, X^c]}_T = v_q^{-1} \sum_{i=1}^{n-q+1} \underbrace{|\Delta X_{t_{n,i}}|^r |\Delta X_{t_{n,i+1}}|^r |\Delta X_{t_{n,i+2}}|^r ... |\Delta X_{t_{n,i+q-1}}|^r}_{\text{q factors}} \tag{20}$$

for integer q where $q \geq 2$, where

$$r = \frac{2}{q}, \quad v_q = (E|N(0,1)|^r)^q,$$

and $\Delta X_{t_{n,i}} = X_{t_{n,i}} - X_{t_{n,i-1}}$.

Obviously, if there are no jumps, $E\widehat{[X^c, X^c]}_T = \sigma^2 T$. More generally, if there are jumps, the estimator will typically remain consistent. In the case where the jumps are from an α-stable process, and if $r < \alpha$, one obtains that $E\widehat{[X^c, X^c]}_T \to \sigma^2 T = [X^c, X^c]_T$ as $n \to \infty$.

How about the error in the estimate? There are two components in the error, the stochastic term and the bias term:

$$\widehat{[X^c, X^c]}_T - \sigma^2 T = \underbrace{\widehat{[X^c, X^c]}_T - E\widehat{[X^c, X^c]}_T}_{\text{stochastic term}} + \underbrace{E\widehat{[X^c, X^c]}_T - \sigma^2 T}_{\text{bias term}}$$

The stochastic part is straightforward. Under the condition that

$$r < \frac{\alpha}{2}, \tag{21}$$

the Central Limit Theorem (CLT) guarantees that the stochastic part of the error

$$\widehat{[X^c, X^c]}_T - E(\widehat{[X^c, X^c]}_T) = O_p(n^{-1/2}), \tag{22}$$

under P. To show this, one can invoke the Lindeberg CLT, see Theorem 27.1 (p. 359-360) of Billingsley (1995). (This shows the order $O_p(n^{-1/2})$ for sums of every q'th term. One can then average over the q possible sums.) The next question is what happens to the non-stochastic part of the error, *i.e.*, the bias.

To find the bias, note that if h_n satisfies the conditions of Theorem 1, the Lindeberg (triangular array) version of the CLT assures that $\widehat{[X^c, X^c]}_T - E_{h_n}(\widehat{[X^c, X^c]}_T) = O_p(n^{-1/2})$ under P_{h_n}, and hence under P by Theorem 1. If this is combined with (22), one obtains $E_{h_n}(\widehat{[X^c, X^c]}_T) - E(\widehat{[X^c, X^c]}_T) = O(n^{-1/2})$. In fact, by making use of the explicit (Edgeworth) form of the likelihood (equations (A.10)-(A.14)), it is clear that $n^{1/2}(\widehat{[X^c, X^c]}_T - E_{h_n}\widehat{[X^c, X^c]}_T)$ is asymptotically independent of the likelihood ratio dP_{h_n}/dP. Thus

$$E_{h_n}(\widehat{[X^c, X^c]}_T) - E(\widehat{[X^c, X^c]}_T) = o(n^{-1/2}). \tag{23}$$

Hence we can calculate the bias $E(\widehat{[X^c, X^c]}_T) - \sigma^2 T$ by calculating $E_{h_n}(\widehat{[X^c, X^c]}_T) - \sigma^2 T$, up to and including the order $O(n^{-1/2})$.

The latter problem (calculating $E_{h_n}(\widehat{[X^c, X^c]}_T) - \sigma^2 T$) is considerably easier than the original problem (calculating $E(\widehat{[X^c, X^c]}_T) - \sigma^2 T$). This is because the latter problem uses the measure P_{h_n}, which only involves a Brownian motion and a compound Poisson process.

The following lemma leads us to study the bias of $\widehat{[X^c, X^c]}_T$.

Lemma 1. *Suppose that $1 < \alpha < 3/2$ and $0 < r < \alpha$. Set $\gamma = 1/(3-\alpha)$, and $h_n = h_0 n^{-\gamma}$. Then*

$$\begin{aligned} &\text{effect of big jumps: } E_{h_n}\widehat{[X^c, X^c]}_T - \sigma^2_{h_n} T = n^{-\frac{2-\alpha}{3-\alpha}} h_0^{2-\alpha} C_1 + o(n^{-1/2}) \text{ and} \\ &\text{effect of small jumps: } \qquad \sigma^2_{h_n} T - \sigma^2 T = n^{-\frac{2-\alpha}{3-\alpha}} h_0^{2-\alpha} C_2 + o(n^{-1/2}), \end{aligned} \tag{24}$$

where C_1 only depends on α and r, and $C_2 = Tc^+(1+k_\alpha)(2-\alpha)^{-1}$. On the other hand, if $0 < \alpha \le 1$ and $0 < r < \alpha$, set $\gamma > 1/2$ and $h_n = h_0 n^{-\gamma}$, then both effects given above are of order $o(n^{-1/2})$.

The effect of small jumps follows directly from (14), and is taken in for comparison. To derive the effect of big jumps, one uses the technology in Theorem 1; the proof is given in Appendix B.

We can now study the bias of the multipower estimator. For $\alpha > 1$ and $r < \alpha/2$, by adding the two effects in Lemma 1, and using formula (23), one obtains that

$$E\widehat{[X^c, X^c]}_T - \sigma^2 T = n^{-\frac{2-\alpha}{3-\alpha}} h_0^{2-\alpha}(C_1 + C_2) + o(n^{-1/2}). \tag{25}$$

However, since the left hand side of (25) is independent of h_0, then the same must hold for the right hand side. In particular, since $\frac{2-\alpha}{3-\alpha} < \frac{1}{2}$, $C_1 + C_2 = 0$. In light of (22), it is proved that the multipower estimator is asymptotically unbiased for $\alpha < 3/2$.

Theorem 2. *Let $\sigma^2 > 0$ and let J_t be an α-stable process. Assume (21). If $0 < \alpha < 3/2$, then $E\widehat{[X^c, X^c]}_T - \sigma^2 T = o(n^{-1/2})$.*

As a corollary, we obtain

Corollary 1. *(CLT for the multipower estimator). Under the conditions of Theorem 2,*

$$n^{1/2}(\widehat{[X^c, X^c]}_T - \sigma^2 T) \xrightarrow{\mathcal{L}} N(0, \gamma^2), \tag{26}$$

where

$$\gamma^2 \;=\; \sigma^4 T^2 \left(a^q + 2a\frac{a^{q-1} - 1}{a - 1} - (2q - 1) \right), \tag{27}$$

and $a = E|N(0,1|^{2r}/(E|N(0,1|^r)^2$.

The corollary follows from Theorem 2 by the same reasoning that was used to show (22). The easiest route is to invoke Theorem 3.2 in Hall and Heyde (1980)). Theorem 2 says that for $0 < \alpha < 3/2$, the bias term is smaller than the stochastic variance term. It should be noted that for $\alpha < 1$ a similar result has been shown by Barndorff-Nielsen, Shephard, and Winkel (2006). The problem gets harder the closer α gets to 2, the current Theorem 2 extends the existing results. It is conjectured that the same technology can be used for all $\alpha \in (0, 2)$, with a more careful treatment. This is beyond the scope of this paper.

Theorem 2 is a surprising and positive result. It is the first time that an estimator of σ^2 has been shown to be asymptotically unbiased and $\sqrt{n}$-consistent for $\alpha > 1$ (α unknown); the α known case has been treated by Aït-Sahalia and Jacod (2007). Obviously, similar results hold when the process is just compound Poisson. Where does this leave the identification problem discussed at the beginning of the paper?

6 When Good Estimators do Bad Things

Section 5 shows that there can exist estimators $\widehat{[X^c, X^c]}_T$ of $[X^c, X^c]_T$ which are $\sqrt{n}$-consistent and asymptotically unbiased in the case when $\alpha > 1$ (unknown α). The identification problem discussed in Section 3 is quite subtle. In this Section one shall see how the bias shows up in estimators of $[X^c, X^c]_T$ through non-uniform convergence of the estimator to the parameter.

Definition 3. *For unknown α, an estimator $\widehat{[X^c, X^c]}_T$ of $[X^c, X^c]_T$ will be called* good *if*
(i) the estimator is $\sqrt{n}$-consistent for all P_h, where h is fixed and $h \geq 0$, and
(ii) the limit theorem for the stochastic part (the Central Limit Theorem or other) applies uniformly: $\widehat{[X^c, X^c]}_T = E_{h_n}\widehat{[X^c, X^c]}_T + O_p(n^{-1/2})$ under P_{h_n} whenever h_n is bounded in n.

Condition (i) includes the possibility of $h = 0$. Since $P_0 = P$, condition (i) guarantees that a good estimator is $\sqrt{n}$-consistent under P. On the other hand, condition (ii) requires a good estimator to have an $O_p(n^{-1/2})$ stochastic term in its estimation error, under all P_{h_n}.

Example 1. *Under the conditions of Theorem 2, the multipower estimator is good. Condition (i) is obviously satisfied for $h \neq 0$, and follows from Corollary 1. Condition (ii) is satisfied in view of the Lindeberg CLT, as used repeatedly in Section 5.*

Example 2. *In the unknown α case, Condition (ii) is satisfied for the Aït-Sahalia and Jacod (2007) estimator. This follows from the construction of the estimator in Section 6 of their paper.*

For estimators in general, it is usually easy to check whether Condition (ii) holds or not. This is because the uniform rate in the CLT follows from the triangular array nature of the Lindeberg Central Limit Theorem (see Theorem 27.2 in Billingsley (1995), and Theorem 3.2 in Hall and Heyde

(1980)). On the other hand, it is the $\sqrt{n}$-consistency under P (part of Condition (i)) that is difficult to verify for most estimators.

Note that under Definition 3, a good estimator satisfies

$$E_{h_n}(\widehat{[X^c, X^c]}_T) - E(\widehat{[X^c, X^c]}_T) = O(n^{-1/2}), \tag{28}$$

by the same development as the one leading to formula (23) in Section 5. (Because of the general assumptions on $\widehat{[X^c, X^c]}_T$, however, we cannot assert that the difference is $o(n^{-1/2})$). However, from (14), one then gets the following.

Theorem 3. *Let $\sigma^2 > 0$ and let J_t be an α-stable process. For unknown $\alpha > 1$ and $h_n = h_0 n^{-1/(3-\alpha)}$, and for a good estimator $\widehat{[X^c, X^c]}_T$, its bias under P_{h_n} is*

$$E_{h_n}(\widehat{[X^c, X^c]}_T) - \sigma^2_{h_n} T = -n^{-\frac{2-\alpha}{3-\alpha}} T c^+ (1 + k_\alpha)(2-\alpha)^{-1} h_0^{2-\alpha} + O(n^{-1/2}). \tag{29}$$

From (i) of Definition 3, the bias of a good estimator $\widehat{[X^c, X^c]}_T$ is no larger than $O(n^{-1/2})$ for any of the probabilities P_h, whether α-stable ($h = 0$) or compound Poisson ($h \neq 0$). The lesson of Theorem 3 is that the asymptotics may not be very reliable for finite n. In fact, for any large but finite n, there is a compound Poisson process (with law P_{h_n}) for which the estimator has a bias that is substantially larger than its standard deviation. This is because $(2-\alpha)/(3-\alpha) < 1/2$.

One description of the phenomenon is that a "good" estimator *shrinks* towards the infinite activity model: it tends to underestimate the continuous-part quadratic variation $[X^c, X^c]_T$ under P_h, and thus overestimate the quadratic variation due to jumps under P_h. The concept of shrinkage goes back to Stein (1956) and James and Stein (1961). Another way to think of this is: for every n, one can always find an even smaller cutoff (h) so that more of the small moves in X are "detected" as jumps, as a consequence, some of continuous evolutions are over-detected under P_h and thus $\widehat{[X^c, X^c]}_T$ has a negative bias.

Theorem 3 shows that a good estimator could behave badly for finite samples. The undesirable property is caused by the non-uniform rate amongP_hs. If the rate is uniform, how fast can it be? This is given in the next theorem, to round off the discussion.

Theorem 4. *Assume the conditions of Theorem 1, and also that $\alpha > 1$, $h_n = h_0 n^{-1/(3-\alpha)}$. Let $\widehat{[X^c, X^c]}_T$ be any estimator of $[X^c, X^c]_T$. Let $\beta > 0$ be such that $\widehat{[X^c, X^c]}_T - [X^c, X^c]_T = O_p(n^{-\beta})$ under both P and P_{h_n}. Then $\beta \leq (2-\alpha)/(3-\alpha)$.*

It is instructive for the understanding of Theorem 1 to read the following proof:

Proof of Theorem 4. Under the assumptions of the theorem, $n^\beta(\widehat{[X^c, X^c]}_T - \sigma^2_{h_n} T)$ is $O_p(1)$ under P_{h_n}. By Theorem 1, $n^\beta(\widehat{[X^c, X^c]}_T - \sigma^2_{h_n} T)$ is also $O_p(1)$ under P. On the other hand, again by assumptions of the theorem, $n^\beta(\widehat{[X^c, X^c]}_T - \sigma^2 T) = O_p(1)$ under P. Thus, $n^\beta(\sigma^2_{h_n} T - \sigma^2 T) = O_p(1)$ under P. Since $\sigma^2_{h_n}$ and σ^2 are nonrandom, $\sigma^2_{h_n} T - \sigma^2 T = O(n^{-\beta})$. Hence, by (14), the theorem is proved.

7 Conclusion

There is an increasing literature on separating the volatility due to continuous evolution from the volatility due to jumps. The separation problem is closely related to the identification of continuous evolution as opposed to small jumps in the process X. The papers argue that the distinction between jumps and continuity is best thought of as pragmatic rather than absolute. For those small jumps whose sizes are below certain threshold, one is better off to identify the "undetectable" jumps with continuous evolution – as the title of the paper suggests, what you don't know cannot hurt you – to achieve easy analysis and computation.

Department of Finance, The University of Illinois at Chicago, Chicago, IL60607, U.S.A.; lanzhang@uic.edu; http://tigger.uic.edu/∼lanzhang/.

REFERENCES

AÏT-SAHALIA, Y. (2002): "Telling from Discrete Data Whether the Underlying Continuous-Time Model is a Diffusion," *Journal of Finance*, 57, 2075–2112.

AÏT-SAHALIA, Y., AND J. JACOD (2004): "Fisher's Information for Discretely Sampled Lévy Processes," Discussion paper, Princeton University and Université de Paris VI.

——— (2007): "Volatility Estimators for Discretely Sampled Lévy Processes," *Annals of Statistics*, 35, 335–392.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2004): "Power and bipower variation with stochastic volatility and jumps (with discussion)," *Journal of Financial Econometrics*, 2, 1–48.

——— (2005): "Impact of jumps on returns and realised variances: Econometric analysis of time-deformed Lévy processes," *Journal of Econometrics, forthcoming*, –, –.

——— (2006): "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics*, 4, 1–30.

BARNDORFF-NIELSEN, O. E., N. SHEPHARD, AND M. WINKEL (2006): "Limit theorems for multipower variation in the presence of jumps," *Stochastic Processes and Applications*, 116, 796–806.

BERTOIN, J. (1998): *Lévy Processes.* Cambridge University Press, Cambridge, UK.

BILLINGSLEY, P. (1995): *Probability and Measure.* Wiley, New York, third edn.

FAN, J., AND Y. WANG (2005): "Multi-scale Jump and Volatility Analysis for High-Frequency Financial Data," Discussion paper, Princeton University.

HÁJEK, J., AND Z. SIDAK (1967): *Theory of Rank Tests.* Academic Press, New York.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application.* Academic Press, Boston.

HUANG, X., AND G. TAUCHEN (2006): "The relative contribution of jumps to total price variance," *Journal of Financial Econometrics*, 4, 456–499.

JACOD, J. (2006): "Asymptotic Properties of Realized Power Variations and related Functionals of Semimartingales," *Stochastic Processes and Applications*, p. (to appear).

JACOD, J., AND P. PROTTER (1998): "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26, 267–307.

JACOD, J., AND A. N. SHIRYAEV (2003): *Limit Theorems for Stochastic Processes.* Springer-Verlag, New York, second edn.

JAMES, W., AND C. STEIN (1961): "Estimation with quadratic loss," *Proc. 4th Berk Symp Math. Statist. Prob.*, 1, 311–319, Univ of Calif. Press, Berkeley, CA.

LeCam, L. (1986): *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

LeCam, L., and G. Yang (1986): *Asymptotics in Statistics: Some Basic Concepts.* Springer-Verlag, New York, second edn.

Lee, S. Y., and P. A. Mykland (2006): "Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics," Discussion paper, Georgia Institute of Technology and The University of Chicago, to appear in *Review of Financial Studies.*

Mancini, C. (2004): "Estimating the Integrated Volatility in Stochastic Volatility Models with Lévy Type Jumps," Discussion paper, Universita di Firenze.

——— (2006): "Non Parametric Threshold Estimation in Models with Stochastic Diffusion Coefficient and Lévy Jumps," Discussion paper, Universita di Firenze.

Protter, P. (2004): *Stochastic Integration and Differential Equations: A New Approach.* Springer-Verlag, New York, second edn.

Stein, C. (1956): "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," *Proc. 3rd Berk Symp Math. Statist. Prob.*, 1, 197–206, Univ of Calif. Press, Berkeley, CA.

APPENDIX: PROOFS

A Proof of Theorem 1

We begin by showing the following auxiliary theorem. Define, under P and P_h,

$$\begin{aligned}(P)\ X_t^{(s)} &= X_0 + \mu t + \sigma W_t + J_t^{(s)}\\ (P_h)\ X_t^{(s)} &= X_0 + \mu_h t + \sigma_h W_t.\end{aligned} \tag{A.1}$$

whence, obviously, $X_{t_i} = X_{t_i}^{(s)} + J_{t_i}^{(l)}$ under both P and P_h.

Theorem 5. *Assume the conditions in Theorem 1. Consider P and P_h as probabilities on $X_{t_{n,i}}^{(s)}$. Let $h_n = h_0 n^{-\gamma}$. Then, if $\gamma < 1/(3-\alpha)$,*

$$\frac{dP_{h_n}}{dP}(X_0^{(s)}, X_{t_{n,1}}^{(s)}, ..., X_T^{(s)}) \to 1 \tag{A.2}$$

in P-probability as $n \to \infty$. Similarly, if $\gamma = 1/(3-\alpha)$ and $\alpha > 1$,

$$\frac{dP_{h_n}}{dP}(X_0^{(s)}, X_{t_{n,1}}^{(s)}, ..., X_T^{(s)}) = O_{ui}(1) \tag{A.3}$$

in P-probability as $n \to \infty$.

Proof of Theorem 5. We here show the statement (A.3). The result (A.2) follows similarly.

Note that an equivalent statement to (A.3) is that

$$\frac{dP}{dP_{h_n}}(X_0^{(s)}, X_{t_{n,1}}^{(s)}, ..., X_T^{(s)}) = O_{ui}(1) \tag{A.4}$$

in P_{h_n}-probability as $n \to \infty$. It is thus suffice to show (A.4). Without loss of generality, assume that $\mu = 0$ and $\sigma = 1$. Let f_n be the density of the distribution of $U_n = (X_\delta^{(s)} - X_0^{(s)})/\sqrt{\delta}$ under P, where $\delta = T/n$. Write $R_n = (J_\delta^{(s)} - J_0^{(s)})/\sqrt{\delta}$, so that $U_n = Z_n + R_n$, where Z_n is $N(0,1)$.

Since J_t is an α-stable process, it has Lévy measure given by (15). By Proposition 1, $J_t^{(s)}$ is a martingale and has moments of all orders. Hence, for $p \geq 2$, the cumulants of R_n under P are then given by

$$\begin{aligned}\mathrm{cum}_p(R_n) &= \delta^{-p/2} E\langle J^{(s)}, ..., J^{(s)}\rangle_\delta\\ &= \delta^{1-p/2}\int_{-G(h_n)}^{h_n} x^p \nu(dx)\\ &= \delta^{1-p/2}(c^+ + (-1)^p c^- k_\alpha^{p-\alpha})(p-\alpha)^{-1} h_n^{p-\alpha}(1+o(1)),\end{aligned} \tag{A.5}$$

where $\langle J^{(s)}, ..., J^{(s)} \rangle_t$ is the pth order predictable variation of $J^{(s)}$. By plugging in $\delta = T/n$ and $h = h_0 n^{-\gamma}$, one therefore obtains

$$\mathrm{cum}_p(R_n) = n^{\frac{p}{2}-1-(p-\alpha)\gamma} T^{1-\frac{p}{2}} (c^+ + (-1)^p c^- k_\alpha^{p-\alpha})(p-\alpha)^{-1} h_0^{p-\alpha}(1+o(1)), \tag{A.6}$$

with $\gamma = (3-\alpha)^{-1}$.

One obtains (ϕ is the standard normal density), by the independence of Z_n and R_n, and since $E(R_n) = 0$,

$$\begin{aligned} f_n(x) &= E\phi(x - R_n) \\ &= \phi(x) + \frac{1}{2}E(R_n^2)\phi''(x) - \frac{1}{3!}E(R_n^3)\phi'''(x) + ..., \end{aligned} \tag{A.7}$$

which is to say that f_n has a valid Edgeworth expansion around ϕ. At the same time, if $g_n(x)$ is the density of $(X_\delta^{(s)} - X_0^{(s)})/\sqrt{\delta}$ under P_c, one gets a similar expansion, but with moments calculated under P_h rather than P. The two first moments match by construction between P and P_h, while all higher order cumulants under P_h are zero. At the same time

$$g_n(x) = \frac{1}{\sigma_{h_n}} \phi\left(\frac{x}{\sigma_{h_n}}\right). \tag{A.8}$$

One therefore obtains further that

$$f_n(x) = g_n(x)\{1 + \frac{1}{3!}\mathrm{cum}_3(R_n)h_{n,3}(x) + \frac{1}{4!}\mathrm{cum}_{n,4}(R_n)h_4(x) + \frac{1}{72}\mathrm{cum}_3(R_n)^2 h_{n,6}(x) + O(n^{-\frac{\alpha}{3-\alpha}})\}, \tag{A.9}$$

since $\alpha < 2$, and where the $h_{n,p}$ are the usual Hermite polynomials, here given by $g_n(x)h_{n,p}(x) = (-1)^p g_n^{(p)}$ (they are defined relative to g_n; the standardized ones are defined relative to ϕ). Note that the $h_{n,p}$'s depend on n. Hence,

$$\begin{aligned} \log f_n(x) = \log g_n(x) \\ &+ \frac{1}{3!}\mathrm{cum}_3(R_n)h_3(x) + \frac{1}{4!}\mathrm{cum}_4(R_n)h_4(x) + \frac{1}{72}\mathrm{cum}_3(R_n)^2(h_6(x) - h_3(x)^2) \\ &+ O(n^{-\frac{\alpha}{3-\alpha}}) + O(n^{-\frac{5-\alpha}{2(3-\alpha)}}). \end{aligned} \tag{A.10}$$

Note that $\mathrm{cum}_3(R_n) = O(n^{-\frac{1}{2}})$ and $\mathrm{cum}_4(R_n) = O(n^{\frac{1}{3-\alpha}})$, so the included terms may (depending on α) be smaller than the neglected terms. This does not alter our argument. Also note that the order $O(n^{-\frac{\alpha}{3-\alpha}})$ relates only to terms that are linear in the Hermite polynomials, while the order $O(n^{-\frac{5-\alpha}{2(3-\alpha)}})$ relates to all other terms.

Since $\int h_p(x)g_n(x)dx = 0$, and by uniform integrability, one therefore obtains that

$$\begin{aligned} E_c\left(\log f_n(U_n) - \log g_n(U_n)\right) &= -\frac{1}{72}\mathrm{cum}_3(R_n)^2 E_c h_3(U_n)^2 + O(n^{-\frac{5-\alpha}{2(3-\alpha)}}) \\ \mathrm{Var}_c\left(\log f_n(U_n) - \log g_n(U_n)\right) &= \frac{1}{36}\mathrm{cum}_3(R_n)^2 E_c h_3(U_n)^2 + O(n^{-\frac{5-\alpha}{2(3-\alpha)}}). \end{aligned} \tag{A.11}$$

Since

$$\begin{aligned}\log \frac{dP}{dP_{h_n}}&(X_0, X_{t_{n,1}}, ..., X_T)\\ &= \sum_{i=1}^{n} \log f_n((X^{(s)}_{\delta i} - X^{(s)}_{\delta(i-1)})/\sqrt{\delta}) - \log g_n((X^{(s)}_{\delta i} - X^{(s)}_{\delta(i-1)})/\sqrt{\delta}),\end{aligned} \tag{A.12}$$

(A.11) implies (A.4) by the (triangular array i.i.d.) Central Limit Theorem. Uniform integrability follows since the limit in law integrates to one and since both the prelimiting and limiting quantities are nonnegative. The theorem is therefore proved.

Proof of Theorem 1. We here show the statement (13). The result (12) follows similarly. The result will be proved if one can show the stronger statement (A.3). This is because

$$\frac{dP_h}{dP}(X^{(s)}_0, X^{(s)}_{t_{n,1}}, ..., X^{(s)}_T, J^{(l)}_0, J^{(l)}_{t_{n,1}}, ..., J^{(l)}_T) = \frac{dP_h}{dP}(X^{(s)}_0, X^{(s)}_{t_{n,1}}, ..., X^{(s)}_T) \tag{A.13}$$

since X^s_t and $J^{(l)}_t$ are independent (see Theorem 39 (p. 29) in Protter (2004)). We therefore obtain that

$$\frac{dP_h}{dP}(X_0, X_{t_{n,1}}, ..., X_T) = E\left(\frac{dP_h}{dP}(X^{(s)}_0, X^{(s)}_{t_{n,1}}, ..., X^{(s)}_T) \mid X_0, X_{t_{n,1}}, ..., X_T\right). \tag{A.14}$$

Let η_n be a sequence which converges in law to η under P. Also, let $\exp(\zeta V - \frac{1}{2}\zeta)$ be the limit in law of $\frac{dP_h}{dP}(X^{(s)}_0, X^{(s)}_{t_{n,1}}, ..., X^{(s)}_T)$ under P. We then obtain from (A.14) and Theorem 5 and its proof, that for any bounded continuous function g,

$$E_{h_n}(g(\eta_n)) \to Eg(\eta)\exp(\zeta V - \frac{1}{2}\zeta^2). \tag{A.15}$$

The result is then shown.

B Proof of Lemma 1

We show that the effect of big jumps is of order $o(n^{-1/2})$ for $0 < \alpha \leq 1$, while for $1 < \alpha < 3/2$,

$$E_{h_n}(|\Delta X_{t_{n,1}}/\sqrt{\Delta t_{n,1}}|^r) - \sigma^r_{h_n} E(|N(0,1)|^r) = n^{-\frac{2-\alpha}{3-\alpha}} h_0^{2-\alpha} C'_1 + o(n^{-1/2}) \tag{B.16}$$

where C'_1 only depends on α and r. The result then follows by Taylor expansion, since $E_h[\widehat{X^c, X^c}]_T = v_q^{-1} T \frac{n-q+1}{n} E_{h_n}(|\Delta X_{t_{n,1}}/\sqrt{\Delta t_{n,1}}|^r)^q$.

Before starting the proof, note that from equation (14), by Taylor expansion,

$$\sigma^v_h = \sigma^v + \frac{v}{2}\sigma^{v-2} n^{-\frac{2-\alpha}{3-\alpha}} h_0^{2-\alpha} c^+ (1 + k_\alpha)(2-\alpha)^{-1} + O(n^{-2\frac{2-\alpha}{3-\alpha}}). \tag{B.17}$$

We now turn to the big jumps. Since, $\alpha < 3/2$, $\lambda_h \Delta t_{n,1} = O(h^{-\alpha} n^{-1}) = O(n^{\alpha\gamma - 1})$. For $\alpha \leq 1$ and $\gamma \approx 1/2$, this goes to zero. Also, for $\alpha > 1$ and $\gamma = 1/(3-\alpha)$, $\alpha\gamma - 1 = \alpha/(3-\alpha) - 1 = (2\alpha - 3)/(3-\alpha)$, which is < 0 if $\alpha < 3/2$.

Next need to compute $E_h(|\Delta X_{t_{n,1}}/\sqrt{\Delta t_{n,1}}|^r)$. Let $Y_1, Y_2, ...$ be iid with distribution as in (19). One obtains, with $h = h_n$,

$$\begin{aligned}
E_h(|\Delta X_{t_{n,1}}/\sqrt{\Delta t_{n,1}}|^r) &= E_h(|N(0,\sigma_h^2) + \left(\sum_{i=1}^{N_{\Delta t_{n,1}}} Y_i\right)/\sqrt{\Delta t_{n,1}}|^r) \\
&= E_h(|N(0,\sigma_h^2)|^r)P_h(N_{\Delta t_{n,1}} = 0) \\
&+ E_h(|N(0,\sigma_h^2) + Y_1/\sqrt{\Delta t_{n,1}}|^r)P_h(N_{\Delta t_{n,1}} = 1) + ... \\
&= E_h(|N(0,\sigma_h^2)|^r)\exp(-\lambda_h \Delta t_{n,1}) \\
&+ E_h(|N(0,\sigma_h^2) + Y_1/\sqrt{\Delta t_{n,1}}|^r)(\lambda_h \Delta t_{n,1})\exp(-\lambda_h \Delta t_{n,1}) + ... \\
&= E_h(|N(0,\sigma_h^2)|^r) + b_n + O((\lambda_h \Delta t_{n,1})^2), \qquad \text{(B.18)}
\end{aligned}$$

where

$$\begin{aligned}
b_n &= (\lambda_h \Delta t_{n,1})\left[E_h(|N(0,\sigma_h^2) + Y_1/\sqrt{\Delta t_{n,i}}|^r) - E_h(|N(0,\sigma_h^2)|^r)\right] \\
&= Tn^{-1}\frac{c^+ + c^- k_\alpha^{-\alpha}}{\alpha}h^{-\alpha}\left[E_1(|N(0,\sigma_h^2) + T^{-1/2}n^{1/2}hY_{1,1}|^r) - E_1(|N(0,\sigma_h^2)|^r)\right], \qquad \text{(B.19)}
\end{aligned}$$

where we have used (18) and that $Y_1 = hY_{1,1}$, in law, where $Y_{1,1}$ is the law obtained by using Q with $h = 1$. Obviously, $E(|Y_{1,1}|^r) < \infty$ for $r < \alpha$. Setting $h = h_n = h_0 n^{-\gamma}$, and recalling that $\gamma > 1/2$, one gets that for $\alpha \le 1$,

$$\begin{aligned}
b_n + O((\lambda_h \Delta t_{n,1})^2) &= o(n^{-1+\alpha\gamma}) \\
&= o(n^{-1/2}). \qquad \text{(B.20)}
\end{aligned}$$

In the case when $\alpha > 1$, set

$$\begin{aligned}
m(x) &= E(|N(0,1)+x|^r) \\
&= \int_{-\infty}^{\infty} |u|^r \phi(u-x)du \\
&= \phi(x)(2\pi)^{1/2} \int_{-\infty}^{\infty} |u|^r \exp(xu)\phi(u)du \\
&= \phi(x)(2\pi)^{1/2} E\left\{|N(0,1)|^r \exp(xN(0,1))\right\} \\
&= \phi(x)(2\pi)^{1/2} \sum_{p=0}^{\infty} \frac{1}{p!} x^p E\left\{|N(0,1)|^r N(0,1)^p\right\} \\
&= \phi(x)(2\pi)^{1/2} \sum_{p\geq 0, 2|p} \frac{1}{p!} x^p 2E\left\{N(0,1)^{p+r} I_{\{N(0,1)>0\}}\right\} \\
&= \phi(x)(2\pi)^{1/2} \sum_{p\geq 0, 2|p} \frac{1}{p!} x^p E\left\{|N(0,1)|^{p+r}\right\} \\
&= \phi(x)(2\pi)^{1/2} \sum_{p=0}^{\infty} \frac{1}{(2p)!} x^{2p} E\left\{|N(0,1)|^{2p+r}\right\} \\
&= \phi(x)(2\pi)^{1/2} \sum_{p=0}^{\infty} \frac{1}{(2p)!} x^{2p} 2^{\frac{r+2p}{2}} \pi^{-\frac{1}{2}} \Gamma(\frac{r+2p+1}{2}) \\
&= \phi(x) \sum_{p=0}^{\infty} \frac{1}{(2p)!} 2^{\frac{r+2p+1}{2}} \Gamma(\frac{r+2p+1}{2}) x^{2p} \qquad \text{(B.21)}
\end{aligned}$$

because

$$E|N(0,1)|^r = 2^{\frac{r}{2}} \pi^{-\frac{1}{2}} \Gamma(\frac{r+1}{2})$$

where Γ is the Gamma-function.

Now define

$$f(a) = E_1[m(aY_{1,1})]. \qquad \text{(B.22)}$$

Note that since $Y_{1,1}$ has density proportional to $|y|^{-\alpha-1}$ in the tails,

$$\sum_{p=0}^{\infty} \frac{1}{(2p)!} 2^{\frac{r+2p+1}{2}} \Gamma(\frac{r+2p+1}{2}) E_1[(aY_{1,1})^{2p}\phi(aY_{1,1})] < \infty \qquad \text{(B.23)}$$

so that $f(a)$ equals the sum in (B.23) by Fubini's Theorem. Similarly, derivatives of f can be obtained by differentiating the sum term by term. (And f is infinitely many times differentiable.)

Note that since f is an even function, $f'(0) = f'''(0) = ... = 0$. Hence,

$$\begin{aligned}
b_n &= Tn^{-1}\frac{c^+ + c^- k_\alpha^{-\alpha}}{\alpha} h^{-\alpha}\sigma_h^r \left[f(\sigma_h^{-1}T^{-1/2}n^{1/2}h) - f(0) \right] \\
&= Tn^{-1}\frac{c^+ + c^- k_\alpha^{-\alpha}}{\alpha} h^{-\alpha}\sigma_h^r \left\{ \left(\sigma_h^{-1}T^{-1/2}n^{1/2}h\right)^2 f''(0) + O((n^{1/2}h)^4) \right\} \\
&= h^{2-\alpha} f''(0) \frac{c^+ + c^- k_\alpha^{-\alpha}}{\alpha}\sigma_h^{r-2} + O(nh^{4-\alpha}) \\
&= n^{-\frac{2-\alpha}{3-\alpha}} h_0^{2-\alpha} f''(0) \frac{c^+ + c^- k_\alpha^{-\alpha}}{\alpha}\sigma_h^{r-2} + O(n^{-\frac{1}{3-\alpha}}) \\
&= n^{-\frac{2-\alpha}{3-\alpha}} h_0^{2-\alpha} f''(0) \frac{c^+ + c^- k_\alpha^{-\alpha}}{\alpha}\sigma^{r-2} + O(n^{-\frac{1}{3-\alpha}}) + O(n^{-2\frac{2-\alpha}{3-\alpha}}) \qquad \text{(B.24)}
\end{aligned}$$

since $h_n = h_0 n^{-\gamma}$, with $\gamma = 1/(3-\alpha)$, and by setting $v = r-2$ in (B.17). Since $\alpha < 3/2$, the remainder terms in (B.27) are of order $o(n^{-1/2})$, in other words,

$$b_n = n^{-\frac{2-\alpha}{3-\alpha}} h_0^{2-\alpha} f''(0) \frac{c^+ + c^- k_\alpha^{-\alpha}}{\alpha}\sigma^{r-2} + o(n^{-1/2}). \qquad \text{(B.25)}$$

It remains to deal with the remainder term $O((\lambda_h \Delta t_{n,1})^2)$ in (B.18). Note that

$$O((\lambda_h \Delta t_{n,1})^2) = O(h^{-2\alpha}n^{-2}) = O(n^{2\frac{2\alpha-3}{3-\alpha}}), \qquad \text{(B.26)}$$

so if $\alpha < 4/3$ is term is of order $o(n^{-1/2})$ without any further arguments. It therefore follows from (B.18) that

$$E_h(|\Delta X_{t_{n,1}}/\sqrt{\Delta t_{n,1}}|^r) = n^{-\frac{2-\alpha}{3-\alpha}} h_0^{2-\alpha} f''(0) \frac{c^+ + c^- k_\alpha^{-\alpha}}{\alpha}\sigma^{r-2} + o(n^{-1/2}). \qquad \text{(B.27)}$$

To show the result (B.27) for $4/3 \le \alpha < 3/2$, one observes that there is a finite integer v_0 so that $(\lambda_h \Delta t_{n,1})^{v_0} = o(n^{-1/2})$. We therefor need to to show that the coefficients in front of $(\lambda_h \Delta t_{n,1})^v$ in (B.18) are zero for $2 \le v < v_0$. To see this for $v = 2$, for example, note that similarly to the argument above, one can write the second term as

$$(\lambda_h \Delta t_{n,1})^2 \left[f_2(\sigma_h^{-1}T^{-1/2}n^{1/2}h) - f_2(0) \right], \qquad \text{(B.28)}$$

where $f(a) = E_1[m(a(Y_{1,1} + Y_{2,1}))]$. This function is also even and infinitely many times differentiable, so that (B.28) is of order $O((\lambda_h \Delta t_{n,1})^2 (n^{1/2}h)^2) = O(n^{-1}h^{2(1-\alpha)}) = o(n^{-1/2})$.

Now (B.27) for $\alpha < 3/2$ is shown, Lemma 1 is then proved.