



HAL
open science

Volatility Estimators for Discretely Sampled Lévy Processes

Yacine Aït-Sahalia, Jean Jacod

► **To cite this version:**

Yacine Aït-Sahalia, Jean Jacod. Volatility Estimators for Discretely Sampled Lévy Processes. 2006. hal-00004898v2

HAL Id: hal-00004898

<https://hal.science/hal-00004898v2>

Preprint submitted on 19 Apr 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

VOLATILITY ESTIMATORS FOR DISCRETELY SAMPLED LEVY PROCESSES

BY YACINE AÏT-SAHALIA¹ AND JEAN JACOD²

Princeton University and Université de Paris-6

This paper provides rate-efficient estimators of the volatility parameter
in the presence of Lévy jumps.

1. Introduction. In this paper, we continue the study started in [2], about the estimation of parameters when one observes a Lévy process X at n regularly spaced times $\Delta_n, 2\Delta_n, \dots, n\Delta_n$, with Δ_n going to 0 as $n \rightarrow \infty$. In our earlier paper, we were concerned with the asymptotic behavior of the Fisher information, with the objective of establishing a benchmark for what efficient estimators are able to achieve in that context. Now, we wish to exhibit estimators which both achieve that rate and can be explicitly computed.

We want to estimate a positive parameter σ , which we call volatility, in the model

$$(1) \quad X_t = \sigma W_t + Y_t,$$

where W is a standard Wiener process or, more generally, a symmetric stable process of index β , and the process Y is another Lévy process without Wiener part and with jumps “dominated” in a sense we make precise below by those of W . Allowing for jumps is of great interest in mathematical finance, in the diverse contexts of option

¹Supported in part by NSF Grant SES-0350772.

²Supported in part by the CNRS.

AMS 2000 subject classifications. Primary 62F12, 62M05; secondary 60H10, 60J60.

Key words and phrases. Diffusions with jumps, efficiency, discrete sampling

pricing, testing for the presence of jumps in asset prices, interest rate modelling, risk management, optimal portfolio choice, stochastic volatility modelling or for the purpose of better describing asset returns data (see the references cited in [2]).

Our aim is to construct estimators for σ which behave under the model (1) “as well as” under the model

$$(2) \quad X_t = \sigma W_t,$$

asymptotically as $\Delta_n \rightarrow 0$ and $n \rightarrow \infty$. This is in line with the results of [2], in which we proved that property for the Fisher information. In other words, we want to be able to estimate the volatility parameter σ at the same rate when Y , a jump perturbation of W , is present as when it is not. In some applications, Y may represent frictions that are due to the mechanics of the trading process, or in the case of compound Poisson jumps it may represent the infrequent arrival of relevant information related to the asset. Given that both W and Y contribute to the overall observed noise in X , it is not a priori obvious that it should be possible to estimate σ equally well (at least in the rate sense) with and without Y . Beyond the robustness to misspecification risk that such a result affords, it also for instance paves the way for risk management or option hedging that is able to target the “ W risk” (continuous when $\beta = 2$) separately from the “ Y risk” (discontinuous).

We distinguish between a *parametric* case, where the law of Y is known, and a *semiparametric* case, where it is not. We show that, in the parametric case, one can find estimators which are asymptotically efficient in the Cramer–Rao sense, meaning that the asymptotic estimation variance is equivalent as $n \rightarrow \infty$ to the inverse of the Fisher information for the model (2) without the perturbation Y . This is possible when the law of Y is completely known. In the semiparametric case, where that law is unknown, obtaining asymptotically efficient estimators requires Δ_n to go fast enough to 0; but we can then exhibit estimators that are efficient uniformly when the law of Y stays in a set sufficiently separated from the law of W . And in general

we can exhibit a large class of estimators which are consistent and achieve a specified rate (although not the efficient rate).

A distinctive feature of the present paper is that we construct estimators which are as simple as possible to implement. For example, in the parametric situation where the law of Y is known, one can in principle compute the MLE, which is of course efficient. In practice, this is hardly feasible, as the likelihood function derived from the convolution of the densities of W and Y will in most situations not be available in closed form. So we provide a number of other – much simpler – estimators which are not as good (in the sense of not reaching the Cramer-Rao lower bound in general) but not too bad either (in the sense of achieving the efficient rate of convergence).

The paper is organized as follows. In Section 2, we specify our estimating setting. Section 3 is devoted to estimating equations: the estimators we propose all fall in that class and we state a general result which covers them all. Sections 4 and 5 are devoted to the parametric and semiparametric cases respectively. Some examples are developed in Section 6, 7, 8 and 9, where we consider specific types of estimating equations such as the empirical characteristic function, power variations and power variations with truncation.

2. The setting. With $X_0 = 0$, we observe n i.i.d. increments from the Lévy process (1),

$$(3) \quad \chi_i^n = X_{i\Delta_n} - X_{(i-1)\Delta_n}.$$

W is a symmetric stable process of index $\beta \in (0, 2]$, characterized by

$$(4) \quad \mathbb{E}(e^{iuW_t}) = e^{-t|u|^\beta/2}$$

so that, when $\beta = 2$, W is a standard Wiener process. The parameter to be estimated is σ , and we will single out two situations concerning the parameter space Θ : either $\Theta = (0, \infty)$, or Θ is a compact subset of $(0, \infty)$.

The law of Y (as a process) is entirely specified by the law G_Δ of the variable Y_Δ for any given $\Delta > 0$. We write $G = G_1$, and we recall that the characteristic function of G_Δ is given by the Lévy-Khintchine formula

$$(5) \quad \mathbb{E}(e^{ivY_\Delta}) = \exp \Delta \left(ivb - \frac{cv^2}{2} + \int F(dx) (e^{ivx} - 1 - ivx1_{\{|x| \leq 1\}}) \right)$$

where (b, c, F) is the “characteristic triple” of G (or, of Y): $b \in \mathbb{R}$ is the drift of Y , and $c \geq 0$ the local variance of the continuous part of Y , and F is the Lévy jump measure of Y , which satisfies $\int (1 \wedge x^2) F(dx) < \infty$. We will denote by $\mathbf{P}_{\sigma, G}$ the law of the process X .

We make Y “dominated” by W in the following sense: G belongs to the class \mathcal{G}_β , defined as follows. Let first Φ be the class of all increasing and bounded functions $\phi : (0, 1] \rightarrow \mathbf{R}_+$ having $\lim_{x \downarrow 0} \phi(x) = 0$. Then we set

$$(6)$$

$\mathcal{G}(\phi, \alpha)$ = the set of all infinitely divisible distributions with $c = 0$ and, for all $x \in (0, 1]$,

$$\text{then } \begin{cases} x^\alpha F([-x, x]^c) \leq \phi(x) & \text{if } \alpha < 2 \\ x^2 F([-x, x]^c) \leq \phi(x) \quad \text{and} \quad \int_{\{|y| \leq x\}} |y|^2 F(dy) \leq \phi(x) & \text{if } \alpha = 2, \end{cases}$$

$$(7) \quad \mathcal{G}'(\phi, \alpha) = \{G \in \mathcal{G}(\phi, \alpha), \quad G \text{ is symmetrical about } 0\},$$

$$(8) \quad \mathcal{G}_\alpha = \cup_{\phi \in \Phi} \mathcal{G}(\phi, \alpha), \quad \mathcal{G}'_\alpha = \cup_{\phi \in \Phi} \mathcal{G}'(\phi, \alpha),$$

and we have

$$(9) \quad \begin{cases} \alpha \in (0, 2] & \Rightarrow \mathcal{G}_\alpha = \{G \text{ is infinitely divisible, } c = 0, \lim_{x \downarrow 0} x^\alpha F([-x, x]^c) = 0\} \\ \alpha = 2 & \Rightarrow \mathcal{G}_2 = \{G \text{ is infinitely divisible, } c = 0\}. \end{cases}$$

Now we recall some results from [2]. The variable W_1 admits a C^∞ density h_β , which is differentiable in the state variable (the derivative is denoted by h'_β). Then we set

$$(10) \quad \check{h}_\beta(w) = h_\beta(w) + wh'_\beta(w), \quad \tilde{h}_\beta(w) = \frac{\check{h}_\beta(w)^2}{h_\beta(w)}, \quad \bar{h}_\beta(w) = \frac{wh'_\beta(w)}{h_\beta(w)}, \quad \mathcal{I}(\beta) = \int \tilde{h}_\beta(w) dw,$$

so in fact $\mathcal{I}(\beta)$ is the Fisher information when we estimate σ on the basis of the single observation σW_1 and for the parameter value $\sigma = 1$. The functions \check{h}_β and \tilde{h}_β and \bar{h}_β are also C^∞ , and satisfy for some constant c_β :

$$(11) \quad \begin{cases} \beta < 2 & \Rightarrow \quad h_\beta(w) + |\check{h}_\beta(w)| + |\tilde{h}_\beta(x)| \leq \frac{c_\beta}{1+|w|^{1+\beta}}, \quad |\bar{h}_\beta(w)| \leq c_\beta, \\ \beta = 2 & \Rightarrow \quad \check{h}_\beta(w) = (1-w^2)h_\beta(w), \quad \tilde{h}_\beta(x) = (1-w^2)^2 h_\beta(w), \quad \bar{h}_\beta(w) = -w^2, \end{cases}$$

and of course $h_2(w) = e^{-w^2/2}/\sqrt{2\pi}$, so in particular $\mathcal{I}(\beta) = 2$.

If we have a single observation X_Δ there is a (finite) Fisher information for estimating σ , which we denote by $I_\Delta(\sigma, G)$. With n observed increments the corresponding Fisher information becomes

$$(12) \quad I_{n, \Delta_n}(\sigma, G) = nI_{\Delta_n}(\sigma, G).$$

The main result of [2], as far as the parameter σ is concerned, is summarized in the following:

Theorem 1. *a) If $G \in \mathcal{G}_\beta$ we have as $\Delta \rightarrow 0$:*

$$(13) \quad I_\Delta(\sigma, G) \rightarrow \frac{1}{\sigma^2} \mathcal{I}(\beta).$$

b) For any $\phi \in \Phi$ we have as $\Delta \rightarrow 0$:

$$(14) \quad \sup_{G \in \mathcal{G}(\phi, \beta)} \left| I_\Delta(\sigma, G) - \frac{\mathcal{I}(\beta)}{\sigma^2} \right| \rightarrow 0.$$

c) For each n let G^n be the standard symmetric stable law of index α_n , with α_n a sequence strictly increasing to β . Then for any sequence $\Delta_n \rightarrow 0$ such that

$(\beta - \alpha_n) \log \Delta_n \rightarrow 0$ (i.e. the rate at which $\Delta_n \rightarrow 0$ is slow enough), the sequence of numbers $I_{\Delta_n}(\sigma, G^n)$ converges to a limit which is strictly less than $\mathcal{I}(\beta)/\sigma^2$.

Part (a) of the above theorem and (12) hint towards the existence of estimators $\hat{\sigma}_n$ such that $\sqrt{n}(\hat{\sigma}_n - \sigma)$ converges to a centered Gaussian variable with variance $\sigma^2/\mathcal{I}(\beta)$ under $\mathbf{P}_{\sigma, G}$, when $G \in \mathcal{G}_\beta$ is known: this is the parametric situation, and we will propose such estimators in Section 4 below. In the semiparametric situation where G is unknown, (c) suggests that we cannot achieve the same rate, unless, as given in (b), we know that G is in the class $\mathcal{G}(\phi, \alpha)$ for some $\alpha < \beta$ and some function $\phi \in \Phi$.

As a matter of fact, we can do slightly better. If $\phi(x) = \zeta > 0$ for all x , we can still define $\mathcal{G}(\phi, \alpha)$ by (6), although ϕ no longer belongs to Φ . We denote such a class by $\overline{\mathcal{G}}(\zeta, \alpha)$, that is we introduce the notation (we do not need to distinguish $\alpha < 2$ and $\alpha = 2$ here):

(15)

$\overline{\mathcal{G}}(\zeta, \alpha)$ = the set of all infinitely divisible distributions with $c = 0$ and, for all $x \in (0, 1]$,
then $x^\alpha F([-x, x]^c) \leq \zeta$,

(16) $\overline{\mathcal{G}}'(\zeta, \alpha) = \{G \in \overline{\mathcal{G}}(\zeta, \alpha), G \text{ is symmetrical about } 0\}$,

(17) $\overline{\mathcal{G}}_\alpha = \cup_{\zeta > 0} \overline{\mathcal{G}}(\zeta, \alpha), \quad \overline{\mathcal{G}}'_\alpha = \cup_{\zeta > 0} \overline{\mathcal{G}}'(\zeta, \alpha)$.

The connection with the previous classes is as follows:

(18) $\mathcal{G}(\phi, \alpha) \subset \overline{\mathcal{G}}(\phi(1), \alpha), \quad \mathcal{G}_\alpha \subset \overline{\mathcal{G}}_\alpha \subset \cap_{\alpha' > \alpha} \mathcal{G}_{\alpha'}, \quad \mathcal{G}_2 = \overline{\mathcal{G}}_2$.

For example, \mathcal{G}_0 is the class of all G 's for which Y is a pure drift ($Y_t = bt$), whereas $\overline{\mathcal{G}}_0$ is the class of all G 's for which Y is a compound Poisson process plus a drift. Also, any stable process Y with index $\alpha < 2$ belongs to $\overline{\mathcal{G}}_\alpha$, but not to \mathcal{G}_α .

3. About estimating equations. The practical estimators we will propose for σ are all obtained by setting an estimating equation (also known as a generalized moment condition) to zero. We prove here a general result about the asymptotic properties of such estimators, which will be used several times below. Similar general results for estimating equations are of course known (see various forms in [5], [6] and [7]), but we adapt them here to our setting with assumptions (by no means minimal) that are sufficient in our context.

Recall that we want to estimate a parameter $\sigma > 0$. At stage n we observe p_n i.i.d. random variables χ_i^n and introduce two auxiliary variables $S_n > 0$ and $Q_n \in \mathbf{R}$. Under the associated probability measure $\mathbf{P}_{n,\sigma}$ we suppose that the families (S_n, Q_n) and $(\chi_i^n : 1 \leq i \leq p_n)$ are independent, and of course $p_n \rightarrow \infty$. Let us introduce the following conditions:

Assumption 1 (A1). *If $\sigma_n \rightarrow \sigma > 0$ then $S_n \rightarrow \sigma$ in \mathbf{P}_{n,σ_n} -probability.*

Assumption 2 (A2). *If $\sigma_n \rightarrow \sigma > 0$ then the sequence $(Q_n | \mathbf{P}_{n,\sigma_n})$ is tight.*

Next we consider two families $(f_{n,s,q})_{s>0}$ and $(H_{n,s})_{s>0,q \in \mathbf{R}}$ of functions on \mathbf{R} and $(0, \infty)$ respectively, to be specified later but with adequate integrability and smoothness properties, and we associate the estimating function

$$(19) \quad U_{n,s,q}(u) = \frac{1}{p_n} \sum_{i=1}^{p_n} (f_{n,s,q}(\chi_i^n) - H_{n,s}(u)).$$

In this exactly-identified context, we set

$$(20) \quad \hat{\sigma}_n(s, q) = \begin{cases} \text{the } u > 0 \text{ with } U_{n,s,q}(u) = 0 \text{ which is closest to } s & \text{if it exists} \\ 1 & \text{otherwise} \end{cases}$$

(if $U_{n,s,q} = 0$ has two closest solutions at equal distance of s , we select the smallest one). We also set

$$(21) \quad F_{n,s,q}(\sigma) = \mathbf{E}_{n,\sigma}(f_{n,s,q}(\chi_i^n)), \quad F_{n,s,q}^{(2)}(\sigma) = \mathbf{E}_{n,\sigma}(f_{n,s,q}(\chi_i^n)^2).$$

Note in particular that we are not assuming that the estimating equation is correctly centered: correct centering would require using $F_{n,s,q}$ instead of $H_{n,s}$. $H_{n,s}$ may be equal to $F_{n,s,q}$, but can also be just an approximation to it (in which case we will talk about “approximate centering”) that may for instance be valid as $n \rightarrow \infty$. Incorrect centering leads to estimators that are asymptotically biased, although that effect can be mitigated as $n \rightarrow \infty$ if $H_{n,s}$ approximates $F_{n,s,q}$ (see Assumption (B5) below).

Let us now list a series of assumptions on the previous functions:

Assumption 3 (B1). *We have $\sup_{n \geq 1, s > 0, q \in \mathbf{R}} \|f_{n,s,q}\|^4/p_n < \infty$, where $\|f\|$ is the sup-norm.*

Assumption 4 (B2). *$H_{n,s}$ is continuously differentiable.*

Assumption 5 (B3). *For all $s > 0$ there is a differentiable function \bar{F}_s on $(0, \infty)$, such that whenever $s_n \rightarrow s$ then H_{n,s_n} and H'_{n,s_n} converge locally uniformly to \bar{F}_s and \bar{F}'_s respectively.*

Assumption 6 (B4). *$\bar{F}'_s(s) \neq 0$ for all $s > 0$.*

Assumption 7 (B5). *$F_{n,s_n,q_n}^{(2)}(u_n)$ converges to a limit $F^{(2)}(u)$ for any two sequences u_n and s_n converging to the same limit $u > 0$ and any bounded sequence q_n .*

Assumption 8 (B6). *There is a sequence $w_n \rightarrow +\infty$ such that $\sup_n w_n |F_{n,s_n,q_n}(u_n) - H_{n,s_n}(u_n)| < \infty$ for any two sequences u_n and s_n converging to the same limit $u > 0$ and any bounded sequence q_n .*

Then we have the following:

Theorem 2. *Assume (A1), (A2) and (B1)–(B6).*

a) The sequence $((w_n \wedge \sqrt{p_n})(\hat{\sigma}_n(S_n, Q_n) - \sigma_n))$ is tight under $\mathbf{P}_{n,\sigma}$, uniformly in n and in σ in any compact subset of $(0, \infty)$.

b) If $w_n/\sqrt{p_n} \rightarrow \infty$, then the sequence $(\sqrt{p_n}(\hat{\sigma}_n(S_n, Q_n) - \sigma_n))$ converges in law under $\mathbf{P}_{n,\sigma}$, uniformly in σ in any compact subset of $(0, \infty)$, towards the centered normal distribution with variance $\Xi^2(\sigma) := (F^{(2)}(\sigma) - \bar{F}_\sigma(\sigma)^2) / \bar{F}'_\sigma(\sigma)^2$.

We devote the remainder of this section to proving this theorem. First, we state a lemma which gathers some classical limit theorems on i.i.d. triangular arrays. For each n let $(\zeta_i^n : i = 1, \dots, p_n)$ be real-valued and i.i.d. random variables, possibly defined on different probability spaces $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$ when n varies. Then:

Lemma 1. *Assume that ζ_i^n is square-integrable, and set $\gamma_n = \mathbf{E}_n(\zeta_i^n)$ and $\Gamma_n = \mathbf{E}_n((\zeta_i^n)^2) - \gamma_n^2$. If $p_n \rightarrow \infty$ and $\Gamma_n/p_n \rightarrow 0$, we have*

$$(22) \quad \frac{1}{p_n} \sum_{i=1}^{p_n} \zeta_i^n - \gamma_n \xrightarrow{\mathcal{L}^2(\mathbf{P}_n)} 0.$$

Furthermore if $\Gamma_n \rightarrow \Gamma$ for some limit $\Gamma > 0$ and if $\mathbf{E}(|\zeta_i^n|^4)/p_n \rightarrow 0$, we have

$$(23) \quad \sqrt{p_n} \left(\frac{1}{p_n} \sum_{i=1}^{p_n} \zeta_i^n - \gamma_n \right) \xrightarrow{\mathcal{L}(\mathbf{P}_n)} \mathcal{N}(0, \Gamma).$$

In the next three lemmas we suppose that $\sigma_n \rightarrow \sigma > 0$, and we write $\mathbf{P}_n = \mathbf{P}_{n,\sigma_n}$.

Lemma 2. *Let $s_n \rightarrow \sigma$ and let q_n be a bounded sequence.*

a) *The sequence $((w_n \wedge \sqrt{p_n}) U_{n,s_n,q_n}(\sigma_n) \mid \mathbf{P}_n)$ is tight.*

b) *If $w_n/\sqrt{p_n} \rightarrow \infty$ then*

$$\sqrt{p_n} U_{n,s_n,q_n}(\sigma_n) \xrightarrow{\mathcal{L}(\mathbf{P}_n)} \mathcal{N}(0, F^{(2)}(\sigma) - \bar{F}_\sigma(\sigma)^2).$$

Proof. We have $U_{n,s_n,q_n}(\sigma_n) = \frac{1}{p_n} \sum_{i=1}^{p_n} \zeta_i^n$, where for each n the ζ_i^n 's are i.i.d. with mean and variance given by

$$\gamma_n = F_{n,s_n,q_n}(\sigma_n) - H_{n,s_n}(\sigma_n), \quad \Gamma_n = F_{n,s_n,q_n}^{(2)}(\sigma_n) - F_{n,s_n,q_n}(\sigma_n)^2,$$

and further $|\zeta_i^n| \leq \alpha_n$ for numbers α_n satisfying $\alpha_n^4/p_n \rightarrow 0$ by (B1). Now (B6) yields that $\gamma_n \rightarrow 0$, hence (B3) yields $F_{n,s_n,q_n}(\sigma_n) \rightarrow \overline{F}_\sigma(\sigma)$. On the other hand, (B5) implies $F_{n,s_n,q_n}^{(2)}(\sigma_n) \rightarrow F^{(2)}(\sigma)$.

Therefore it follows from (23) that

$$(24) \quad \sqrt{p_n} (U_{n,s_n,q_n}(\sigma_n) - \gamma_n) \xrightarrow{\mathcal{L}(\mathbf{P}_n)} \mathcal{N}(0, F^{(2)}(\sigma) - \overline{F}_\sigma(\sigma)^2),$$

and since $\sup_n w_n |\gamma_n| < \infty$ by (B6), we readily get the two results. \square

Lemma 3. a) *The sequence $((w_n \wedge \sqrt{p_n}) U_{n,S_n,Q_n}(\sigma_n) \mid \mathbf{P}_n)$ is tight.*

b) *If $w_n/\sqrt{p_n} \rightarrow \infty$, the sequence $(\sqrt{p_n} U_{n,S_n,Q_n}(\sigma_n) \mid \mathbf{P}_n)$ converges in law towards the centered normal distribution with variance $F^{(2)}(\sigma) - \overline{F}_\sigma(\sigma)^2$.*

Proof. a) Let $V(n, s, q) = (w_n \wedge \sqrt{p_n}) U_{n,s,q}(\sigma_n)$. The previous lemma implies that as soon as the deterministic sequence s_n converges to σ , we have for all $B > 0$:

$$(25) \quad \lim_{A \rightarrow \infty} \sup_{n \geq 1} u_{A,B}(n, s_n) = 0, \quad \text{where } u_{A,B}(n, s) = \sup_{|q| \leq B} \mathbf{P}_n(|V(n, s, q)| > A).$$

If the sequence $(V(n, S_n, Q_n) \mid \mathbf{P}_n)$ is not tight, there exists an infinite sequence n_k such that $\mathbf{P}_{n_k}(|V(n_k, S_{n_k}, Q_{n_k})| > A) \geq 1/A$ for some $A > 0$ and, up to taking a further subsequence still denoted by n_k we can assume by (A1) that $S_{n_k} \rightarrow \sigma$ pointwise. Since (S_n, Q_n) is independent of the family $(V(n, s, q); s > 0, q \in \mathbf{R})$, we get

$$\mathbf{P}_{n_k}(|V(n_k, S_{n_k}, Q_{n_k})| > A) \leq \mathbf{P}_{n_k}(|Q_{n_k}| > B) + \mathbf{E}_{n_k}(u_{A,B}(n_k, S_{n_k})).$$

Then (25) and Lebesgue's Theorem imply that

$$\limsup_k \mathbf{P}_{n_k}(|V(n_k, S_{n_k}, Q_{n_k})| > A) \leq \sup_n \mathbf{P}_n(|Q_n| > B)$$

for all $B > 0$ and, in view of (A2), we deduce that $\limsup_k \mathbf{P}_{n_k}(|V(n_k, S_{n_k}, Q_{n_k})| > A) = 0$: this contradicts the definition of the sequence n_k , and we have the result.

b) Let us denote by V a variable with law $\nu = \mathcal{N}(0, F^{(2)}(\sigma) - \overline{F}_\sigma(\sigma))$. Let $\nu_{n,s,q}$ be the law of $V(n, s, q) := \sqrt{p_n} U_{n,s,q}(\sigma_n)$. The claim amounts to proving that, for all bounded continuous functions g , we have

$$(26) \quad \mathbf{E}_n(g(V(n, S_n, Q_n))) \rightarrow \mathbf{E}(g(V)).$$

For this, it is enough to prove that from any subsequence one can extract a further subsequence along which (26) holds. So, in view of (A1) and (A2) it is no restriction to assume that in fact (S_n, Q_n) converges in law to (σ, Q) for some variable Q .

In fact, due to the independence of (S_n, Q_n) and $(W'(n, s, q) : s > 0, q \in \mathbf{R})$, we can replace the pair (S_n, Q_n) in the left side of (26) by any other pair (S'_n, Q'_n) having the same law than (S_n, Q_n) and still independent of $(W'(n, s, q) : s > 0, q \in \mathbf{R})$. Therefore, using the Skorokhod representation theorem, we can indeed assume that (S_n, Q_n) converges pointwise to (σ, Q) . Then

$$\mathbf{E}_n(g(V(n, S_n, Q_n))) = \mathbf{E}_n\left(\int \nu_{n,S_n,Q_n}(dx)g(x)\right).$$

Since $S_n \rightarrow \sigma$ and $Q_n \rightarrow Q$, one deduces from Lemma 2–(b) that the sequence $\int \nu_{n,S_n,Q_n}(dx)g(x)$ converges pointwise to $\int \nu(dx)g(x) = \mathbf{E}(g(V))$, and it is bounded by $\|g\|$, so Lebesgue's Theorem yields (26). \square

Lemma 4. *The sequence $\hat{\sigma}_n$ converges in \mathbf{P}_n -probability to σ .*

Proof. Exactly as in the previous proof, without loss of generality we can assume that the pair (S_n, Q_n) converges pointwise to (σ, Q) with Q a suitable random variable.

Lemma 3 implies that $U_{n,S_n,Q_n}(\sigma_n) \rightarrow 0$ in probability (recall that both w_n and p_n go to infinity). Observe that

$$U_{n,S_n,Q_n}(u) - U_{n,S_n,Q_n}(\sigma_n) = H_{n,S_n}(\sigma_n) - H_{n,S_n}(u),$$

which by (B3) converges (pointwise) locally uniformly in u towards $H(u) := \overline{F}_\sigma(\sigma) - \overline{F}_\sigma(u)$. Hence $U_{n,S_n,Q_n}(u)$ also converges locally uniformly in u towards $H(u)$, in \mathbf{P}_n -probability. But by (B4) the function H is null at σ and is either strictly decreasing

or strictly increasing in a neighborhood of σ : then the definition (20) of $\widehat{\sigma}_n(S_n, Q_n)$ immediately gives the result. \square

Finally, we prove Theorem 2:

Proof of Theorem 2. As usual, to get the local uniformity in σ for the tightness in (a) or the convergence in (b), it is enough to obtain the tightness (resp. convergence) under $P_n = P_{n, \sigma_n}$ for any sequence $\sigma_n \rightarrow \sigma > 0$. Let us write for simplicity $\widehat{\sigma}_n = \widehat{\sigma}_n(S_n, Q_n)$ and $U_n = U_{n, S_n, Q_n}$.

By (B2), U_n is continuously differentiable. We deduce from Lemma 4 the existence of sets A_n with $P_n(A_n) \rightarrow 1$, such that on A_n we have $U'_n(\widehat{\sigma}_n) = 0$, and thus Taylor's formula yields a random variable T_n taking its values between σ_n and $\widehat{\sigma}_n$, and such that

$$(27) \quad U_n(\sigma_n) = -(\widehat{\sigma}_n - \sigma_n)U'_n(T_n) \quad \text{on the set } A_n.$$

Observe that $U'_n(T_n) = -H'_{n, S_n}(T_n)$. Since both S_n and T_n converge in probability to σ , (B3) implies that $U'_n(T_n) \rightarrow -\overline{F}'_\sigma(\sigma)$ in probability. Since $\overline{F}'_\sigma(\sigma) \neq 0$ by (B4), all the results of our theorem are now easily deduced from (27) and Lemma 3. \square

With this general result in hand, we now turn to our specific situation: estimating σ in the presence of the Lévy process Y , first when the law of Y is known and second when it is not.

4. Estimation of σ in the parametric case. In this section, we study the estimation of σ when the law of Y , i.e., the measure $G \in G_\beta$, is known. We will construct a class of estimating equations for σ , with χ_i^n given by (3).

4.1. *Construction of the estimators.* In the sequel the number $\beta \in (0, 2]$ is fixed and does not usually appear explicitly in our notation. A constant which depends only on β and on another parameter γ is denoted by C_γ , and it may change from line to line. If $G \in G_\alpha$ with $\alpha \leq \beta$, and with the associated process Y , we set

$$(28) \quad b'(G, \alpha) = \begin{cases} b - \int_{\{|x| \leq 1\}} xF(dx) & \text{if } \alpha < 1 \\ b & \text{if } \alpha \geq 1, \end{cases} \quad Z_\Delta(\alpha) := \Delta^{-1/\beta} (Y_\Delta - b'(G, \alpha)\Delta)$$

and we let $G'_{\Delta, \alpha}$ denote the law of $Z_\Delta(\alpha)$. Then we define the “modified increments” (recall (3)):

$$(29) \quad \chi_i^n(G) = \Delta_n^{-1/\beta} (\chi_i^n - b'(G, \beta)\Delta_n).$$

Next, for any $\alpha \in (0, 2]$ and any $\phi \in \Phi$ we set for $x \in (0, 1)$:

$$(30) \quad \phi_\alpha(x) = \begin{cases} \frac{\phi(x)}{1-\alpha} & \text{if } \alpha < 1 \\ \phi(x) + \frac{\phi(x)}{\sqrt{\log(1/x)}} + \phi\left(1 \wedge e^{-\sqrt{\log(1/x)}}\right) & \text{if } \alpha = 1 \\ \phi(x) + \frac{\phi(\sqrt{x})}{\alpha-1} + \frac{\phi(1)}{\alpha-1} x^{\frac{\alpha-1}{2}} & \text{if } \alpha > 1. \end{cases}$$

This defines an increasing function $\phi_\alpha : (0, 1] \rightarrow R_+$ having $\phi \leq \phi_\alpha$ and $\phi_\alpha(x) \rightarrow 0$ as $x \rightarrow 0$.

Next, if $G \in G_\alpha$ for some $\alpha \leq \beta$, and $u > 0$ and $v \geq 0$ and $z \in R$ and if k is a bounded function, we set

$$(31) \quad \Psi_{G, \Delta, \alpha, k}(u, v, z) = \int h_\beta(x) dx \int G'_{\Delta, \alpha}(dw) k(ux + vw + z).$$

Finally, we introduce the “tail function”

$$(32) \quad \psi(u) = \mathbf{P}(|W_1| > 1/u) = 2 \int_{1/u}^{\infty} h_\beta(x) dx$$

for $u > 0$ (this depends on β): it is C^∞ , strictly increasing from 0 to 1, with non-vanishing first derivative. So its reciprocal function ψ^{-1} , from $(0, 1)$ into $(0, \infty)$, is also C^∞ and strictly increasing.

Recall that we work here under the assumption that $G \in G_\beta$ is known, and so in particular we know $b'(G, \beta)$; we also have $G \in G(\phi, \beta)$ for some $\phi \in \Phi$. We need first a preliminary estimator, which is constructed as follows. We choose an arbitrary sequence m_n of integers satisfying

$$(33) \quad m_n \uparrow \infty, \quad \frac{m_n}{n} \rightarrow 0$$

and, recalling (29) and (32), we set

$$(34) \quad V_n(G) = \frac{1}{m_n} \sum_{i=1}^{m_n} 1_{\{|\chi_i^{m_n}(G)| > 1\}}, \quad S_n(G) = \begin{cases} \psi^{-1}(V_n(G)) & \text{if } 0 < V_n(G) < 1 \\ 1 & \text{otherwise.} \end{cases}$$

To form an estimating equation for the construction of the final estimator of σ , we choose a function k satisfying

$$(35) \quad \sup_x \frac{|k(x)|}{1 + |x|^\gamma} < \infty, \quad I(k) := \int \check{h}_\beta(x) k(x) dx \neq 0,$$

where the number γ satisfies

$$(36) \quad \gamma \geq 0, \quad \beta \leq 2 \quad \Rightarrow \quad \gamma < \frac{\beta}{2}.$$

Then we set

$$(37) \quad k_n(x) = \begin{cases} k(x) & \text{if } k \text{ is bounded} \\ k(x) 1_{\{|k(x)| \leq \nu_n\}} & \text{otherwise,} \end{cases}$$

where ν_n be an increasing sequence of numbers satisfying

$$(38) \quad \nu_n \rightarrow \infty, \quad \nu_n^2 \phi_\beta(\Delta_n^{1/\beta}) \rightarrow 0, \quad \frac{\nu_n^4}{n} \rightarrow 0,$$

and where ϕ_β is associated with ϕ (a function such that $G \in G(\phi, \beta)$) by (30). Then, with the notation $p_n = n - m_n$, and since each k_n is bounded, we can define the following estimation functions (for $u > 0$):

$$(39) \quad U_{n,G,\phi,k}(u) = \frac{1}{p_n} \sum_{i=m_n+1}^n k_n \left(\frac{\chi_i^{m_n}(G)}{S_n(G)} \right) - \Psi_{G,\Delta_n,\beta,k_n} \left(\frac{u}{S_n(G)}, \frac{1}{S_n(G)}, 0 \right).$$

Finally the estimators for σ are:

$$(40) \quad \hat{\sigma}_n(G, \phi, k) = \begin{cases} \text{the } u > 0 \text{ with } U_{n,G,\phi,k}(u) = 0 \text{ which is closest to } S_n(G) & \text{if it exists} \\ 1 & \text{otherwise.} \end{cases}$$

As the notation suggests, this estimator depend on G and on k in an obvious way, and it depends on ϕ through the choice for k_n made in (38). It also depends on β , but we leave this dependency implicit to avoid cluttering the notation.

4.2. *Asymptotic distribution in the parametric case.* With the function k as in (35), the following defines two finite numbers:

$$(41) \quad J(k) = \mathbf{E}(k(W_1)^2) - (\mathbf{E}(k(W_1)))^2, \quad \Sigma^2(k) = \frac{J(k)}{I(k)^2}.$$

Theorem 3. *Let $\phi \in \Phi$, and let k be a function satisfying (35) for some γ having (36). Suppose also that $\Delta_n \rightarrow 0$.*

a) *The sequence $\sqrt{n} (\hat{\sigma}_n(G, \phi, k) - \sigma)$ converges in law to $N(0, \sigma^2 \Sigma^2(k))$, under $P_{\sigma,G}$, uniformly in $G \in G(\phi, \beta)$ and in $\sigma \in [\varepsilon, 1/\varepsilon]$ for any $\varepsilon > 0$.*

b) *We have $\Sigma^2(k) \geq 1/I(\beta)$, and this inequality is an equality if we choose $k = \bar{h}_\beta$.*

Now we give a number of comments and examples.

Remark 1. *In light of (41), it is of course possible / advisable to select the function k to minimize $\Sigma^2(k)$. The choice $k = \bar{h}_\beta$ is indeed possible: by (11) the function $k = \bar{h}_\beta$ satisfies (35) with $\gamma = 0$ (resp. $\gamma = 2$) if $\beta < 2$ (resp. $\beta = 2$). Such a choice gives asymptotically efficient estimators, in the strong sense that they behave asymptotically like the efficient estimators for the model $X_t = \sigma W_t$ (with no perturbing term Y).*

Remark 2. *To put these estimators in use we would need to numerically compute the function $\Psi_{G,\Delta,\beta,k}(u, v, 0)$, for a single value of v (either 1 or $1/S_n(G)$), and all values of u (in principle). Except in special situations (see for instance Section 6), there is*

no closed form for this function, and we have to resort to numerical integration or to Monte–Carlo techniques. For this it is of course helpful to have a closed form for k (or rather for the truncated k_n). In general, this is not the case for the function $k = \bar{h}_\beta$ (the optimal choice), unless $\beta = 2$.

Remark 3. As an example of function k , we can take $k(x) = |x|^r$, for some $r > 0$ when $\beta = 2$ and $r \in (0, \beta/2)$ otherwise (when $\beta = 2$ and $r = 2$ this is the optimal choice since $\bar{h}_2(x) = -x^2$): the function $\Psi_{G, \Delta_n, \beta, k_n}$ is still not explicit, but it is easily approximated by Monte–Carlo techniques, at last when Y_t can be simulated, or it may be available in closed form for some common distributions of Y . We will do that in some detail in Section 7. In any event, the limiting variance is easy to compute from (41).

Remark 4. Another possibility is to use the empirical characteristic function of the sampled increments, which leads to an closed form expression for $\Psi_{G, \Delta_n, \beta, k_n}$. This will be done in Section 6.

4.3. *Some preliminaries.* Here we gather some results from [2], and also about the functions of (31), which will be used to obtain the previous theorem and for further results as well. First we recall Lemma 2 of [2]: for any $\phi \in \Phi$, and with the notation (30), we have for $\Delta \leq 1$ and $\alpha \leq \beta$ and $K \geq 0$ and some constant $C = C_\alpha$ depending on α only,

$$(42) \quad G \in \mathcal{G}(\phi, \alpha), \quad |g(x)| \leq K(1 \wedge |x|) \quad \implies \quad \mathbf{E}(|g(Z_\Delta(\alpha))|) \leq CK \Delta^{\frac{2(\beta-\alpha)}{\beta(2+\alpha)}} \phi_\alpha(\Delta^{\frac{2+\beta}{\beta(2+\alpha)}}).$$

In fact the proof of this result also works when $\phi(x) = \zeta$ for all x (with ϕ_α substituted with a constant), thus giving

$$(43) \quad G \in \bar{\mathcal{G}}(\zeta, \alpha), \quad |g(x)| \leq K(1 \wedge |x|) \quad \implies \quad \mathbf{E}(|g(Z_\Delta(\alpha))|) \leq CK \zeta \Delta^{\frac{2(\beta-\alpha)}{\beta(2+\alpha)}}.$$

This is not enough for our purposes, at least in the semiparametric situation, and we will need also the next lemma about symmetrical measures:

Lemma 5. *If $\Delta \leq 1$ and $\alpha \leq \beta$ and $K \geq 0$, we have for some constant C depending on α only:*

$$(44) \quad G \in \overline{\mathcal{G}}'(\zeta, \alpha), \quad |g(x)| \leq K(1 \wedge |x|^2) \quad \implies \quad \mathbf{E}(|g(Z_\Delta(\alpha))|) \leq CK\zeta\Delta^{\frac{\beta-\alpha}{\beta}}.$$

Proof. It is similar to the proof of Lemma 2 of [2]. Taking $\eta > 0$, we set $Y_t'' = \sum_{s \leq t} \Delta Y_s 1_{\{|\Delta Y_s| > \eta\}}$ and $Y' = Y - Y''$ and if $G \in \overline{\mathcal{G}}'(\zeta, \alpha)$ then Y is symmetrical and thus we have (47) of the afore-mentioned proof (with ϕ_α substituted with a constant proportional to ζ), that is

$$\mathbf{E}(|Y'_\Delta|^2) \leq C\zeta\Delta\eta^{2-\alpha}$$

for a constant C depending on α only. We also have $Z_\Delta(\alpha) = \Delta^{-1/\beta}Y_\Delta$, hence $|g(Z_\Delta(\alpha))| \leq K\Delta^{-2/\beta}|Y'_\Delta|^2$ on the set $\{Y''_\Delta = 0\}$, whose probability is smaller than $C\zeta\Delta/\eta^\alpha$. Since $|g| \leq K$, we deduce

$$\mathbf{E}(|g(Z_\Delta(\alpha))|) \leq CK\zeta \left(\Delta\eta^{-\alpha} + \Delta^{1-2/\beta}\eta^{2-\alpha} \right).$$

Then take $\eta = \Delta^{1/\beta}$ to obtain the result. □

Next, as soon as the function k satisfies the first half of (35) with some $\gamma \geq 0$ which has $\gamma < \beta$ whenever $\beta < 2$, we set for $u > 0$ and $z \in R$:

$$(45) \quad \Psi_k(u, z) = \int h_\beta(x)k(ux+z) dx = \frac{1}{u} \int h_\beta\left(\frac{x}{u}\right) k(x+z) dx = \frac{1}{u} \int h_\beta\left(\frac{x-z}{u}\right) k(x) dx.$$

(so $\Psi_k(u, z) = \Psi_{G, \Delta, \alpha, k}(u, 0, z)$, which depends neither on G , nor on Δ , nor on α).

Lemma 6. *a) Let k satisfy the first half of (35) with some $\gamma \geq 0$ which has $\gamma < \beta$ whenever $\beta < 2$. Then Ψ_k is C^∞ on $(0, \infty) \times R$. If further $\gamma > 0$ and $\nu \in (0, \infty)$ and*

$k_\nu(x) = k(x)1_{\{|k(x)| \leq \nu\}}$, then for all $K > 0$ there exists $M_{K,k}$ such that

(46)

$$|z| \leq K, \quad \nu \geq M_{K,k} \quad \implies$$

$$\left| \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_k(u, z) - \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_{k_\nu}(u, z) \right| \leq \begin{cases} C_{j,l,k,K} u^{\beta-j} \nu^{1-(l+\beta)/\gamma} & \text{if } \beta < 2 \\ C_{j,l,k,K} u^{j+l+\gamma-1} e^{-\nu^{1/\gamma}/u} & \text{if } \beta = 2. \end{cases}$$

b) If k is bounded, then for all $\eta \in (0, 1)$ we have

$$(47) \quad \eta \leq u \leq 1/\eta \quad \implies \quad \left| \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_k(u, z) \right| \leq C_{l,j,\eta} \|k\|.$$

Proof. (a) If $l \in N$, the j th derivative of $u \mapsto (-1)^l h_\beta^{(l)}(x/u)/u^{l+1}$ takes the form $h_{l,j}(x/u)/u^{j+l+1}$ for a function $h_{l,j}$ satisfying

$$(48) \quad |h_{l,j}(x)| \leq \begin{cases} C_{j,l}/(1 + |x|^{1+l+\beta}) & \text{if } \beta < 2 \\ C_{j,l}(1 + |x|^{2j+2l}) e^{-x^2/2} & \text{if } \beta = 2. \end{cases}$$

In particular the estimate for $\beta < 2$ above also holds for $\beta = 2$, and further $h_{l,j}$ is differentiable and, for all $\beta \in (0, 2]$,

$$(49) \quad |h'_{l,j}(x)| \leq \frac{C_{j,l}}{1 + |x|^{2+l+\beta}}.$$

Therefore we easily deduce from (45) that Ψ_k is C^∞ , with (by differentiating l times the last term in (45), then j times the analogue of the third term with $h_\beta^{(l)}$ instead of h_β):

$$(50) \quad \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_k(u, z) = \frac{1}{u^{j+l+1}} \int h_{l,j}(x/u) k(x+z) dx = \frac{1}{u^{j+l}} \int h_{l,j}(x) k(ux+z) dx.$$

In particular, for some $\varepsilon_k > 0$ depending on the function k , we have

$$\begin{aligned} \left| \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_k(u, z) - \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_{k_\nu}(u, z) \right| &\leq \frac{1}{u^{j+l+1}} \int |k(x+z) - k_\nu(x+z)| h_{l,j}(x/u) dx \\ &\leq \frac{C_k}{u^{j+l+1}} \int_{\{1+|x+z|^\gamma > \nu \varepsilon_k\}} (1 + |x+z|^\gamma) h_{l,j}(x/u) dx. \end{aligned}$$

Then a simple computation, using (48), gives us (46).

(b) When k is bounded, (48) and (50) immediately yield (47). \square

Finally we give estimates for the difference $\Psi_{G,\Delta,\alpha,k}$ and Ψ_k .

Lemma 7. *If k is a bounded function, $\Psi_{G,\Delta,\alpha,k}(u, v, z)$ is C^∞ in (u, z) , and for any $\eta \in (0, 1)$ we have*

$$(51) \quad \eta \leq u \leq 1/\eta \implies \left| \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_{G,\Delta,\alpha,k}(u, v, z) \right| \leq C_{l,j,\eta} \frac{\|k\|}{1 + |z|^{l+\beta}}.$$

Moreover, for all $\eta \in (0, 1)$ we have the following, for all $\Delta \leq 1$ and $z \in \mathbb{R}$ and $u \in [\eta, 1/\eta]$ and $v \in (0, 1/\eta]$:

(i) *If $G \in G(\phi, \alpha)$ (resp. $G \in \overline{\mathcal{G}}(\zeta, \alpha)$), then with ϕ_α given by (30) (resp. $\phi_\alpha \equiv \zeta$):*

$$(52) \quad \left| \frac{\partial^j}{\partial u^j} \Psi_{G,\Delta,\alpha,k}(u, v, z) - \frac{\partial^j}{\partial u^j} \Psi_k(u, 0) \right| \leq C_{j,\eta} \|k\| \left(|z| + \Delta^{\frac{2(\beta-\alpha)}{\beta(2+\alpha)}} \phi_\alpha(\Delta^{\frac{2+\beta}{\beta(2+\alpha)}}) \right),$$

(ii) *If $G \in \overline{\mathcal{G}}'(\zeta, \alpha)$, then*

$$(53) \quad \left| \frac{\partial^j}{\partial u^j} \Psi_{G,\Delta,\alpha,k}(u, v, z) - \frac{\partial^j}{\partial u^j} \Psi_k(u, 0) \right| \leq C_{j,\eta} \|k\| \left(|z| + \zeta \Delta^{\frac{\beta-\alpha}{\beta}} \right),$$

Proof. Observe that $\Psi_{G,\Delta,\alpha,k}(u, v, z) = \int G'_{\Delta,\alpha}(dw) \Psi_k(u, vw + z)$. Then by (47), $\Psi_{G,\Delta,\alpha,k}$ is C^∞ in (u, z) , with

$$(54) \quad \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_{G,\Delta,\alpha,k}(u, v, z) = \int G'_{\Delta,\alpha}(dw) \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_k(u, vw + z),$$

and for any $\eta \in (0, 1)$ we have (51).

Next we prove (i). (49) yields

$$(55) \quad |y| \leq 1 \implies |h_{0,j}(x+y) - h_{0,j}(x)| \leq C_{j,m} \frac{|y|}{1 + |x|^{2+\beta}}.$$

Recalling (50) and (54), we have

$$(56) \quad \frac{\partial^j}{\partial u^j} \Psi_{G,\Delta,\alpha,k}(u, v, z) - \frac{\partial^j}{\partial u^j} \Psi_k(u, z) = \int G'_{\Delta,\alpha}(dw) g(w),$$

where

$$\begin{aligned} g(w) &= \frac{\partial^j}{\partial u^j} \Psi_k(u, vw + z) - \frac{\partial^j}{\partial u^j} \Psi_k(u, z) \\ &= \frac{1}{u^j} \int h_{0,j}(x) (k(ux + vw + z) - k(ux + z)) dx \\ &= \frac{1}{u^j} \int \left(h_{0,j} \left(x - \frac{vw}{u} \right) - h_{0,j}(x) \right) k(ux + z) dx, \end{aligned}$$

for u, v, z, j fixed. Let $\eta \in (0, 1)$, and suppose that $\eta \leq u \leq 1/\eta$ and that $v \leq 1/\eta$. If $|w| \leq 1$ (55) obviously yields $|g(w)| \leq C_{j,\eta} \|k\| |w|$, whereas (47) yields $|g(w)| \leq C_{j,\eta} \|k\|$ always: so we have $|g(w)| \leq C_{j,\eta} \|k\| (|w| \wedge 1)$, and in view of (56) we readily deduce from (42) if $G \in G(\phi, \alpha)$ and (43) if $G \in \overline{\mathcal{G}}(\zeta, \alpha)$ (then $\phi_\alpha \equiv \zeta$), then

$$(57) \quad \left| \frac{\partial^j}{\partial u^j} \Psi_{G,\Delta,\alpha,k}(u, v, z) - \frac{\partial^j}{\partial u^j} \Psi_k(u, z) \right| \leq C_{j,\eta} \|k\| \Delta^{\frac{2(\beta-\alpha)}{\beta(2+\alpha)}} \phi_\alpha(\Delta^{\frac{2+\beta}{\beta(2+\alpha)}}).$$

Moreover (47) yields $\left| \frac{\partial^j}{\partial u^j} \Psi_k(u, z) - \frac{\partial^j}{\partial u^j} \Psi_k(u, 0) \right| \leq C_{j,\eta} \|k\| |z|$, so putting all these together gives (52).

Finally we prove (ii). The function $h_{0,j}$ is C^∞ and all its derivatives satisfy the estimates (48), and in particular $H(x) = \sup_{y \in [x-1/\eta^2, x+1/\eta^2]} |h''_{0,j}(y)|$ is integrable, as well as $h'_{0,j}$. Now we have

$$(58) \quad |w| \leq 1 \quad \Rightarrow \quad \left| h_{0,j} \left(x - \frac{vw}{u} \right) - h_{0,j}(x) - h'_{0,j}(x) \frac{vw}{u} \right| \leq C_{j,\eta} w^2 H(x)$$

as soon as $v < 1/\eta$ and $\eta \leq u \leq 1/\eta$. Therefore we can write $g = g_1 + g_2$, where

$$g_1(w) = \frac{vw}{u^{j+1}} 1_{\{|w| \leq 1\}} \int h'_{0,j}(x) k(ux + z) dx,$$

$$g_2(w) = g(w) 1_{\{|w| > 1\}} + 1_{\{|w| \leq 1\}} \int \left(h_{0,j} \left(x - \frac{vw}{u} \right) - h_{0,j}(x) - h'_{0,j}(x) \frac{vw}{u} \right) k(ux + z) dx.$$

On the one hand, if $G \in \overline{\mathcal{G}}'(\zeta, \alpha)$ then $G'_{\Delta,\alpha}$ is symmetrical about 0, hence $\int g_1(w) G'_{\Delta,\alpha}(dw) = 0$ because g_2 is bounded and odd. On the other hand, (58) plus the integrability of H and the fact that $|g(w)| \leq C_{j,\eta} \|k\|$ yield $|g_2(w)| \leq C_{j,\eta} \|k\| (w^2 \wedge 1)$. Hence, using Lemma 5 we get instead of (57) that

$$(59) \quad \left| \frac{\partial^j}{\partial u^j} \Psi_{G,\Delta,\alpha,k}(u, v, z) - \frac{\partial^j}{\partial u^j} \Psi_k(u, z) \right| \leq C_{j,\eta} \|k\| \zeta \Delta^{\frac{\beta-\alpha}{\beta}},$$

and we conclude (53) as previously. \square

4.4. *Proof of Theorem 3.* We start by proving (b). With the notation $H = \check{h}_\beta/h_\beta$, we observe that in addition to (41), we have

$$I(k) = \mathbf{E}(k(W_1)H(W_1)), \quad \mathcal{I}(\beta) = \mathbf{E}(H(W_1)^2).$$

An integration by parts yields $E(H(W_1)) = 0$, so $J(k) = E(k'(W_1)^2)$ and $I(k) = E(k'(W_1)H(W_1))$ if $k'(x) = k(x) - E(k(W_1))$. The desired inequality, which is $I(k)^2 \leq J(k)I(\beta)$, follows from the Cauchy–Schwarz inequality. If $k = \bar{h}_\beta$ we also have $k = 1 + H$, so this inequality is obviously an equality.

For (a), and since $p_n \sim n$, we apply Theorem 2–(b) with χ_i^n given by (3) and thus $P_{n,\sigma} = P_{\sigma,G}$. The first step consists in proving (A1) for $S_n = S_n(G)$. This amounts to the following lemma, where $\sigma_n \rightarrow \sigma > 0$ and $P_n = P_{\sigma_n,G}$:

Lemma 8. *The sequence S_n converges to σ in probability.*

Proof. By (42) the variables $Z_{\Delta_n}^n(\beta)$ associated with the law G^n converge in law to 0 (because $\phi_\beta(x) \rightarrow$ as $x \rightarrow 0$). The variables χ_i^m , which equal $\sigma_n W_1 + Z_{\Delta_n}^n(\beta)$ in law, converge in law to σW_1 . Hence $\gamma_n := P_n(|\chi_i^m| > 1) \rightarrow \psi(\sigma)$. If $\zeta_i^n = 1_{\{|\chi_i^m| > 1\}}$, (22) applied with $q_n = m_n$ yields $V_n \xrightarrow{\mathbf{P}_{\sigma,G}} \psi(\sigma)$. Since ψ^{-1} is C^∞ and strictly monotone, the result readily follows. \square

Next we set $Q_n = 0$, so (A2) is satisfied, and

$$f_{n,s,q}(x) = k_n \left(\frac{\Delta_n^{-1/\beta}(x - b'(G, \beta)\Delta_n)}{s} \right), \quad H_{n,s}(u) = \Psi_{G, \Delta_n, \beta, k_n} \left(\frac{u}{s}, \frac{1}{s}, 0 \right).$$

Upon comparing (39) and (40) with (19) and (20), we see that $\hat{\sigma}_n(G, \phi, k) = \hat{\sigma}_n(S_n, Q_n)$. Therefore it remains to prove (B1)–(B6) with a sequence w_n satisfying $w_n/\sqrt{p_n} \rightarrow \infty$, and that

$$(60) \quad \Xi^2(\sigma) = \sigma^2 J(k)/I(k)^2.$$

Observe that under $P_{\sigma,G}$ the variables χ_i^n have the same law as $\sigma W_1 + Z_{\Delta_n}(\beta)$. Then (21) gives $F_{n,s,q}(\sigma) = H_{n,s}(\sigma)$. It follows that (B6) holds with w_n arbitrarily large, while (B2) follows from (54).

If k is bounded, hence $k_n = k$, we have $\|f_{n,s}\| \leq \|k\|$ and (B1) is obvious; further, (52) with $\alpha = \beta$ and k^r yields

$$j = 0, 1, \quad r = 1, 2, \quad \eta \leq u \leq \frac{1}{\eta}, \quad v \leq \frac{1}{\eta} \quad \implies$$

$$\left| \frac{\partial^j}{\partial u^j} \Psi_{G,\Delta_n,\beta,k^r}(u, v, 0) - \frac{\partial^j}{\partial u^j} \Psi_{k^r}(u, 0) \right| \leq C_{\eta,k} \phi_\beta(\Delta_n^{1/\beta}),$$

which gives (B3) with $\bar{F}_s(u) = \Psi_k(u/s, 0)$ and (B5) with $F^{(2)}(u) = \Psi_{k^2}(1, 0)$. On the other hand when k is unbounded we have $\|f_{n,s}\| \leq \nu_n$ and thus (B1) follows from (38); further, $\nu_n \rightarrow \infty$ and we can combine (52) with (46) to get for all n large enough:

$$j = 0, 1, \quad r = 1, 2, \quad \eta \leq u \leq \frac{1}{\eta}, \quad v \leq \frac{1}{\eta} \quad \implies$$

$$\left| \frac{\partial^j}{\partial u^j} \Psi_{G,\Delta_n,\beta,k_n^r}(u, v, 0) - \frac{\partial^j}{\partial u^j} \Psi_{k^r}(u, 0) \right| \leq \begin{cases} C_{\eta,k} \left(\nu_n^r \phi_\beta(\Delta_n^{1/\beta}) + \frac{1}{\nu_n^{\beta/r\gamma-1}} \right) & \text{if } \beta < 2 \\ C_{\eta,k} \left(\nu_n^r \phi_2(\Delta_n^{1/2}) + e^{-\eta \nu_n^{1/r\gamma}} \right) & \text{if } \beta = 2. \end{cases}$$

Then, in view of (38) and $2\gamma < \beta$ when $\beta < 2$, we again deduce (B3) with $F_s(u) = \Psi_k(u/s, 0)$ and (B5) with $F^{(2)}(u) = \Psi_{k^2}(1, 0)$.

Since $h_{0,1} = -\check{h}_\beta$, we deduce that $\bar{F}'_\sigma(\sigma) = \Psi'_k(1, 0)/\sigma = -I(k)/\sigma$ (recall (50) and the second part of (35)), hence (B4) holds. We also have $\bar{F}_\sigma(\sigma) = \Psi_k(1, 0) = E(k(W_1))$ and $F^{(2)}(\sigma) = E(k(W_1)^2)$, hence $J(k) = F^{(2)}(\sigma) - F_\sigma(\sigma)^2$ and (60) follows.

5. Estimation of σ in the semiparametric case. Perhaps more realistic than the situation of Theorem 3 is the case where we want to estimate σ , but the measure G is unknown, although we know that it belongs to the class G_β . This is a semiparametric situation: parametric as far as σW_t is concerned, but nonparametric as far as Y_t is concerned. Because G is unknown, the estimating equations in this case must be based on the law of W alone. The challenge is then to achieve rate efficiency despite the lack of information about G .

5.1. *Construction of the estimators.* As said before, we cannot hope for estimators $\hat{\sigma}_n$ that behave nicely for all $G \in G_\beta$ at once. Therefore we suppose that G is unknown, but is known to belong to $\bar{\mathcal{G}}(\zeta, \alpha)$ for some $\alpha < \beta$ and some $\zeta > 0$: we refer to this as Case 1. We also consider a more restrictive situation, called Case 2, for which G is known to belong to the set $\bar{\mathcal{G}}'(\zeta, \alpha)$.

The construction looks pretty much like the previous one, except that besides our preliminary estimator for σ we need to produce an estimator B_n for the drift $b'(G, \alpha)$ in order to remove it. In Case 2, since we know that $b'(G, \alpha) = 0$ we just set

$$(61) \quad B_n = 0.$$

In Case 1 we set $m_n = [\delta n]$ for some arbitrary $\delta \in (0, 1/2)$ ($[x]$ denotes the integer part of x), so that $m_n \sim \delta n$. Then we pick a C^∞ and strictly increasing and odd function θ , with bounded derivative and $\theta(0) = 0$ and $\theta(\pm\infty) = \pm 1$ (for example $\theta(x) = \frac{2}{\pi} \arctan(x)$), and set for $u \in R$

$$(62) \quad R_n(u) = \frac{1}{m_n} \sum_{i=1}^{m_n} \theta(\Delta_n^{-1/\beta}(\chi_i^n - u)).$$

Since $u \mapsto R_n(u)$ is continuous and decreases strictly from $+1$ to -1 as u goes from $-\infty$ to $+\infty$, we can set

$$(63) \quad B_n = \inf(u : R_n(u) = 0) \quad (= \text{the only root of } R_n(\cdot) = 0).$$

Next we construct our preliminary estimator for σ . In Case 1, and with m_n as above, we set $q_n = m_n$ and $p_n = n - 2m_n$. In Case 2, we choose a sequence m_n satisfying (33) and then we set $q_n = 0$ and $p_n = n - m_n$. Then in both cases we set

$$(64) \quad V_n = \frac{1}{m_n} \sum_{i=q_n+1}^{q_n+m_n} 1_{\{|\Delta_n^{-1/\beta}(\chi_i^n - B_n)| > 1\}}$$

and

$$(65) \quad S_n = \begin{cases} \psi^{-1}(V_n) & \text{if } 0 < V_n < 1 \\ 1 & \text{otherwise.} \end{cases}$$

To form estimating equations for σ , we choose a function k satisfying (35) with $\gamma = 0$ (that is, k is bounded and $I(k) \neq 0$). With Ψ_k given by (45) we define the estimating functions (for $u > 0$)

$$(66) \quad U_n(u) = \frac{1}{p_n} \sum_{i=q_n+m_n+1}^n k \left(\frac{\Delta_n^{-1/\beta} (X_i^n - B_n)}{S_n} \right) - \Psi_k \left(\frac{u}{S_n}, 0 \right),$$

and the final estimators

$$(67) \quad \hat{\sigma}_n(k) = \begin{cases} \text{the } u \text{ with } U_n(u) = 0 \text{ which is closest to } S_n & \text{if it exists} \\ 1 & \text{otherwise.} \end{cases}$$

Note that, unlike the centering $\Psi_{G, \Delta_n, \beta, k_n} \left(\frac{u}{S_n(G)}, \frac{1}{S_n(G)}, 0 \right)$ utilized in the parametric case (recall (39)), the centering we now use, based on $\Psi_k \left(\frac{u}{S_n}, 0 \right)$ in (66) does not involve the measure G . Indeed, these estimators depend explicitly on β and k , but on nothing else, and in particular not on G . Observe that they are much easier to compute than the estimator of the parametric case. This is particularly true when $k(x) = \cos(wx)$ for some $w > 0$, since then $\Psi_k(u, 0) = e^{-w^\beta u^\beta / 2}$ is invertible in u , and we will detail this example in the next section, but it is also true in general: first because they depend only on the function $\Psi_k(u, \cdot)$ which is much simpler than the function $\Psi_{G, \Delta, \beta, k}$ accruing in the estimation in the parametric case, second because as a rule $u \mapsto \Psi_k(u, 0)$ is at least “locally invertible” around $u = 1$.

The estimators (66) have formally the same expression in both Case 1 and Case 2, but the preliminary estimators B_n and S_n disagree for the two cases and also $p_n \sim (1 - 2\delta)n$ in Case 1 and $p_n \sim n$ in Case 2, a difference which is important for the asymptotic variance of the estimators. So we will write “the Case 1 version” or “the Case 2 version” of the estimator.

5.2. *Asymptotic distribution in the semiparametric case.* Recall the notation $I(k)$ and $J(k)$ and $\Sigma^2(k)$ of (35) and (41), and let us add some other:

$$(68) \quad \rho(\alpha, \beta) = \frac{2(\beta - \alpha)}{\beta(2 + \alpha)}, \quad \rho'(\alpha, \beta) = \frac{\beta - \alpha}{\beta}.$$

Observe that $\rho(\alpha, \beta) < \rho'(\alpha, \beta)$ always.

Theorem 4. *Let $\alpha \in (0, \beta)$ and $\zeta > 0$, and k be a bounded function with $I(k) \neq 0$, and $\varepsilon \in (0, 1)$. Take the Case 1 version of the estimators.*

a) *If*

$$(69) \quad \sup_n n \Delta_n^{2\rho(\alpha, \beta)} \rightarrow 0,$$

the sequence $\sqrt{n} (\hat{\sigma}_n(k) - \sigma)$ converges in law to $N(0, \sigma^2 \Sigma^2(k)/(1 - 2\delta))$ under $P_{\sigma, G}$, uniformly in $n \geq 1$ and in $\sigma \in [\varepsilon, 1/\varepsilon]$ and in $G \in \bar{\mathcal{G}}(\zeta, \alpha)$.

b) In general, the variables $(\sqrt{n} \wedge \Delta_n^{-\rho(\alpha, \beta)})(\hat{\sigma}_n(k) - \sigma)$ are tight under $P_{\sigma, G}$, uniformly in $\sigma \in [\varepsilon, 1/\varepsilon]$ and in $G \in \bar{\mathcal{G}}(\zeta, \alpha)$ and n .

Theorem 5. *Let $\alpha \in (0, \beta)$ and $\zeta > 0$, and k be a bounded function with $I(k) \neq 0$, and $\varepsilon \in (0, 1)$. Take the Case 2 version of the estimators.*

a) *If*

$$(70) \quad \sup_n n \Delta_n^{2\rho'(\alpha, \beta)} \rightarrow 0,$$

the sequence $\sqrt{n} (\hat{\sigma}_n(k) - \sigma)$ converges in law to $N(0, \sigma^2 \Sigma^2(k))$ under $P_{\sigma, G}$, uniformly in $n \geq 1$ and in $\sigma \in [\varepsilon, 1/\varepsilon]$ and in $G \in \bar{\mathcal{G}}'(\zeta, \alpha)$.

b) In general, the variables $(\sqrt{n} \wedge \Delta_n^{-\rho'(\alpha, \beta)})(\hat{\sigma}_n(k) - \sigma)$ are tight under $P_{\sigma, G}$, uniformly in $\sigma \in [\varepsilon, 1/\varepsilon]$ and in $G \in \bar{\mathcal{G}}'(\zeta, \alpha)$ and n .

The optimal choice of the function k has been discussed after Theorem 3: when $\beta < 2$, we have asymptotic efficiency in the situation of the second theorem above, provided we take $k = \bar{h}_\beta$, and despite the fact that we are in a semiparametric setting. When $\beta = 2$ the choice $k = \bar{h}_\beta$, that is $k(x) = -x^2$, is not permitted in the above theorem, but with $k(x) = -x^2 1_{\{|x| \leq A\}}$ one achieves an asymptotic variance which approaches the optimal variance when A goes to infinity: see Section 7.

Also, some other comments are in order here:

Remark 5. *When α increases, then $\rho(\alpha, \beta)$ and $\rho'(\alpha, \beta)$ decrease, so (69) and (70) are more difficult to obtain and the “rate” in (b) of the two theorems above gets worse, as it should be.*

Remark 6. *In connection with what precedes, one should mention that when (69) fails the actual rate of convergence (that is, a sequence δ_n such that the law of $\delta_n((\hat{\sigma}_n(k) - \sigma)$ converges to a non-degenerate limit, or at least admits among its weak limiting measures a non-degenerate one) is not only unknown, but actually depends on the true underlying (unknown) measure G and in particular on the minimal index α' such that $G \in \overline{\mathcal{G}}_{\alpha'}$ (we know that $\alpha' \leq \alpha$, but the inequality could be strict). In other words, the rate could be for example \sqrt{n} for a particular G , even without (69).*

Remark 7. *However we will see in the examples below (see Section 9 in particular) that (70) is necessary for having convergence to a centered distribution with rate \sqrt{n} and also that the rate in (b) of Theorem 5 is sharp, if we want to have a result which holds uniformly in $G \in \overline{\mathcal{G}}'(\zeta, \alpha)$. We do not know whether (69) or the rate in (b) are optimal for Theorem 4.*

Remark 8. *Of course it might exist other – thoroughly different – estimators behaving better than the $\hat{\sigma}_n(k)$ ’s, and perhaps having a better rate than in (b) of these theorems (the rate cannot be improved in (a), of course). We think this doubtful, however.*

Remark 9. *The most interesting situation is when we have asymptotic efficiency (this happens when G is symmetrical), or at least “rate-efficiency” (that is of order \sqrt{n}). We have this under (69) or (70), which mean that Δ_n goes to 0 fast enough. Of course having $\Delta_n = o(1/n)$ is of no practical use. When $\Delta_n = 1/n$, then rate-efficiency is satisfied as soon as $\alpha \leq 2\beta/(4+\beta)$ for the first theorem and $\alpha \leq \beta/2$ for the second one. If Y is a compound Poisson process with drift, rate efficiency holds as soon as $n\Delta_n^2$ is bounded, whatever $\beta \in (0, 2]$ is (take $\alpha = 0$).*

Remark 10. *When we do not know that G is symmetrical we cannot achieve asymptotic efficiency even under (69). However the asymptotic variances in the two theorems above are the same, up to the factor $1 - 2\delta$: hence by choosing δ small one can approach asymptotic efficiency as much as one wants to.*

5.3. *Proof of Theorems 4 and 5.* As above, we refer to Theorem 4 as to Case 1, and to Theorem 5 as to Case 2. The proof goes through several steps.

1) We fix $\alpha \in (0, \beta)$ and $\zeta > 0$. The sequence Δ_n is fixed, and we set

$$(71) \quad \rho = \begin{cases} \rho(\alpha, \beta) & \text{in Case 1} \\ \rho'(\alpha, \beta) & \text{in Case 2,} \end{cases} \quad \lambda_n = \sqrt{n} \wedge \frac{1}{\Delta_n^\rho}.$$

In order to get tightness or convergence, “uniform” in σ and in G is the relevant class, it is of course enough to take a sequence $\sigma_n \rightarrow \sigma > 0$ and a sequence G^n in $\bar{\mathcal{G}}(\zeta, \alpha)$ (resp. $\bar{\mathcal{G}}'(\zeta, \alpha)$), and to prove the tightness or convergence in law of the normalized estimation errors $\hat{\sigma}_n - \sigma_n$, under the measures $P_n = P_{\sigma_n, G^n}$. Below we fix the sequences σ_n and G^n .

Finally, we denote by $Z_n := Z_{\Delta_n}^n(\alpha)$ the variable associated with the measure G^n by (28), and we set $b'_n = \Delta_n^{1-1/\beta} b'(G^n, \alpha)$, which vanishes in Case 2.

2) Let $Q_n = \lambda_n B'_n$, where $B'_n = (\Delta_n^{-1/\beta} B_n - b'_n)$. We want to prove that the sequence Q_n satisfies (A2). This is obvious in Case 2 because $Q_n = 0$. So we suppose that we are in Case 1. Let us introduce some notation: with $j = 1, 2$ and θ' being the derivative of θ , we put

$$(72) \quad \Gamma_j(\sigma) = \mathbf{E}(\theta(\sigma W_1)^j), \quad \Gamma'_1(\sigma) = \mathbf{E}(\theta'(\sigma W_1))$$

(Γ'_1 is of course the derivative of Γ_1).

Observe that B'_n is the only root of $R_n(\cdot) = 0$, where

$$\mathcal{R}_n(u) = R_n(\Delta_n^{1/\beta}(u + b'_n)) = \frac{1}{m_n} \sum_{i=1}^{m_n} \zeta_i^n(u), \quad \text{with } \zeta_i^n(u) = \theta(\Delta_n^{-1/\beta} \chi_i^n - u - b'_n).$$

The $\zeta_i^n(u)$'s for $i \geq 1$ are i.i.d. with the same law (under P_n) than the variable $\theta(\sigma_n W_1 + Z_n - u)$ (we have used here the scaling property of W).

The functions $\gamma_{n,j}(u) = E_n((\zeta_i^n(u))^j)$, for $j \in N$, are C^∞ and bounded as well as their derivatives, uniformly in u and n , and we can interchange derivation and expectation. So we can apply (43) to the functions $g_{n,j,p}(w) = \int h_\beta(x)(\partial^p \theta^j / \partial u^p)(\sigma_n x + w - u) - (\partial^p \theta^j / \partial u^p)(\sigma_n x - u) dx$, to get for $p, j \in N$:

$$(73) \quad \left| \frac{\partial^p}{\partial u^p} \gamma_{n,j}(u) - \Gamma_{j,p}(\sigma_n, u) \right| \leq C_{p,j} \zeta \Delta_n^\rho, \quad \text{where } \Gamma_{j,p}(v, u) = (-1)^p \int \frac{\partial^p \theta^j}{\partial u^p}(vx - u) h_\beta(x) dx.$$

In particular $\Gamma_{j,0}(\sigma, 0) = \Gamma_j(\sigma)$ for $j = 1, 2$ and $\Gamma_{1,1}(\sigma_n, 0) = \Gamma'_1(\sigma)$ with the notation (72).

Now, R_n also is C^∞ , bounded as well as all its derivatives, uniformly in n , u and ω . So an application of Lemma 1 and the continuity of the functions $\Gamma_{j,p}$ readily yield

$$(74) \quad \frac{\partial^p}{\partial u^p} \mathcal{R}_n(u) \rightarrow \Gamma_{1,p}(\sigma, u) \text{ locally uniformly in } u, \text{ in } \mathbf{P}_n\text{-probability,}$$

$$(75) \quad \eta_n := \sqrt{m_n} (\mathcal{R}_n(0) - \gamma_{n,1}(0)) \xrightarrow{\mathcal{L}(\mathbf{P}_n)} \mathcal{N}(0, \Gamma_2(\sigma) - \Gamma_1(\sigma)^2).$$

The properties of θ imply that $u \mapsto \Gamma_{1,0}(\sigma, \cdot)$ decreases strictly and vanishes at 0; since by construction $R_n(B'_n) = 0$, we deduce from (74) for $p = 0$ that $B'_n \xrightarrow{\mathbf{P}_n} 0$. Another application of (74) yields that $R'_n(B''_n) \xrightarrow{\mathbf{P}_n} \Gamma'_1(\sigma)$ for any sequence B''_n of random variable going to 0 in P_n -probability. Since $R_n(B'_n) = 0$ we have

$$(76) \quad \mathcal{R}'_n(B''_n) B'_n = -\mathcal{R}_n(0) = -\frac{\eta_n}{\sqrt{m_n}} - \gamma_{n,1}(0)$$

for some random variable B_n'' satisfying $|B_n''| \leq |B_n'|$. Moreover $\Gamma_{1,0}(0) = 0$, due to the fact that θ is odd, hence $|\gamma_{n,1}(0)| \leq C\zeta\Delta_n^\rho$ by (73). Since $R_n'(B_n'') \xrightarrow{\mathbf{P}_n} \Gamma_1'(\sigma) \neq 0$, we deduce that $Q_n = \lambda_n B_n'$ satisfies (A2) from (75) (recall $m_n \sim \delta n$ here and (71)).

3) Now we proceed to proving the consistency of the preliminary estimators S_n . In Case 2 the variables V_n and S_n are the variables $V_n(G^m)$ and $S_n(G^m)$ of (29) and (34) (they do not depend on G^m in fact), so the result follows from Lemma 8. In Case 1, set

$$V_n(v) = \frac{1}{m_n} \sum_{i=q_n+1}^{q_n+m_n} 1_{\{|\Delta_n^{-1/\beta}(\chi_i^n - v)| > 1\}}, \quad \delta_n(v) = \mathbf{P}_n(|\Delta_n^{-1/\beta}(\chi_i^n - v)| > 1).$$

Then (22) yields

$$(77) \quad V_n(v_n) - \delta_n(v_n) \xrightarrow{\mathbf{P}_n} 0.$$

However, $\Delta_n^{-1/\beta}(\xi_i^n - v_n)$ has the same distribution as $\sigma_n W_1 + Z_n + b_n' - \Delta_n^{-1/\beta} v_n$, which by (42) converges in law to σW_1 as soon as $b_n' - \Delta_n^{-1/\beta} v_n \rightarrow 0$. Since B_n and $(V_n(v) : v \in R)$ are independent and $B_n' = \Delta_n^{-1/\beta} B_n - b_n' \xrightarrow{\mathbf{P}_n} 0$ because $Q_n = \lambda_n B_n'$ satisfies (A2) and $\lambda_n \rightarrow \infty$, we deduce from (77) that $V_n = V_n(B_n) \xrightarrow{\mathbf{P}_n} \psi(\sigma)$. Then the consistency is proved like in the end of Lemma 8.

4) At this stage we will apply Theorem 2, with the variables (S_n, Q_n) as above and the i.i.d. variables $(\chi_{q_n+m_n+i}^n : 1 \leq i \leq p_n)$. Observe that with the notation (20) and (67), we have $\widehat{\sigma}'_n(k) = \widehat{\sigma}_n(S_n, Q_n)$. We have shown (A1) and (A2) in the two previous steps. Set

$$f_{n,s,q}(x) = k \left(\frac{\Delta_n^{-1/\beta} x - b_n' - q/\lambda_n}{s} \right), \quad H_{n,s}(u) = \Psi_k \left(\frac{u}{s}, 0 \right).$$

Then (21) gives for $r = 1, 2$:

$$F_{n,s,q}(u) = \Psi_{G^n, \Delta_n, \alpha, k} \left(\frac{u}{s}, \frac{1}{s}, -\frac{q}{s\lambda_n} \right), \quad F_{n,s,q}^{(2)}(u) = \Psi_{G^n, \Delta_n, \alpha, k^2} \left(\frac{u}{s}, \frac{1}{s}, -\frac{q}{s\lambda_n} \right).$$

Let us check (B1)–(B6). Since k is bounded, (B1) is obvious, whereas (B2) follows from Lemma 6. Next, if we set $\overline{F}_s(u) = \Psi_k(u/s, 0)$ and $F^{(2)}(u) = \Psi_{k^2}(1, 0)$, Lemma 7 yields for $j = 0, 1$ and $\eta \in (0, 1)$ and $s, u \in [\eta, 1/\eta]$ and $|q| \leq 1/\eta$:

$$\begin{aligned} \left| \frac{\partial^j}{\partial u^j} H_{n,s}(u) - \frac{\partial^j}{\partial u^j} \overline{F}_s(u) \right| &\leq C_{k,\eta} \zeta \Delta_n^\rho, \\ \left| F_{n,s,q}^{(2)}(u) - F_s^{(2)}(u) \right| &\leq C_{k,\eta} \left(\zeta \Delta_n^\rho + \frac{1}{\lambda_n} \right), \\ |F_{n,s,q}(u) - H_{n,s}(u)| &\leq C_{k,\eta} \left(\zeta \Delta_n^\rho + \frac{1}{\lambda_n} \right). \end{aligned}$$

These give (B3) and (B5), and also (B6) with $w_n = \lambda_n$. Finally (B4) holds because $\overline{F}'_s(s) = \psi'_k(1, 0)/s = -I(k)/s$, and (60) holds here as well as in the previous section.

We can thus apply Theorem 2: the sequence $\lambda_n(\hat{\sigma}_n - \sigma_n)$ is tight under P_n in all cases, and this gives the two claims (b). Under (69) or (70) we have $\lambda_n/\sqrt{n} \rightarrow \infty$, hence $\lambda_n\sqrt{p_n} \rightarrow \infty$ as well, so $\sqrt{p_n}(\hat{\sigma}_n - \sigma_n)$ converges in law under P_n to a centered Gaussian variable with variance

$$\Xi^2(\sigma) = \frac{F^{(2)}(\sigma) - \overline{F}_\sigma(\sigma)^2}{\overline{F}_\sigma(\sigma)^2},$$

which in view of $\overline{F}_\sigma(\sigma)^2 = J(k)/\sigma^2$ equals $\sigma^2 \Sigma^2(k)$: since $p_n \sim (1 - 2\delta)n$ in Case 1 and $p_n \sim n$ in Case 2, we obtain the two claims (a).

6. Example: The empirical characteristic function. We now turn to specific estimators. To each specification of an admissible function k (in the sense of satisfying the assumptions of the above results), corresponds an estimator for σ . For instance, one way of estimating a parameter for i.i.d. variables X_j is to use the empirical characteristic function, that is $\sum_{j \in J} \exp(iwX_j)$ for some given w (or several w 's at once) and where J is the index set. If the X_j 's are symmetrical, one should in fact look at the real part only, that is $\sum_{j \in J} \cos(wX_j)$. Other estimators based on the empirical characteristic function in related contexts are given by e.g., [9], [3], [4], Chapter 4 in [11] and [10].

In the parametric situation, at stage n the variable X_j is $\chi_j^n(G)$ and $J = \{m_n + 1, \dots, n\}$. Those variables are “almost” symmetrical (the leading term W coming in them is symmetrical). So we consider for any given $w > 0$ the variable

$$(78) \quad V_n(w) = \frac{1}{p_n} \sum_{i=m_n+1}^n \cos\left(\frac{w\chi_i^n(G)}{S_n(G)}\right),$$

where $S_n(G)$ is the preliminary estimator. In other words, if we take $k(x) = \cos(wx)$ (a bounded function, so $k_n = k$ in (37)), the estimating function of (39) is

$$(79) \quad U_{n,G,\beta,k}(u) = V_n(w) - \Psi_{G,\Delta_n,\beta,k}\left(\frac{u}{S_n(G)}, \frac{1}{S_n(G)}, 0\right).$$

Furthermore, this class of functions k is one for which the function $\Psi_{G,\Delta,\beta,k}$ is explicit, at least when the exponent in the Lévy–Khintchine formula for Y is explicitly known. More precisely, let us write $\rho(u)$ for the exponent in (5), and recall that $E(\exp iuY_t) = \exp t\rho(u)$. Then obviously when $g(x) = e^{ix}$ we have

$$\Psi_{G,\Delta,\beta,g}(u, v, 0) = \exp\left(-\frac{w^\beta u^\beta}{2} + \Delta\rho(wv\Delta^{-1/\beta}) - i w v b'(G, \alpha)\Delta^{1-1/\alpha}\right).$$

Taking the real part, and using (28) and the fact that $G \in G_\beta$, we see that for $k(x) = \cos(wx)$ we have

$$(80) \quad \Psi_{G,\Delta,\beta,k}(u, v, 0) = e^{A_\Delta(u,v)} \cos(B_\Delta(u, v)),$$

where

$$(81) \quad A_\Delta(u, v) = -\frac{w^\beta u^\beta}{2} + \int F(dx) \left(\cos(wv\Delta^{1-1/\beta}x) - 1\right),$$

$$(82) \quad B_\Delta(u, v) = \begin{cases} \int F(dx) \sin(wv\Delta^{1-1/\beta}x), & \text{if } \beta < 1 \\ \int F(dx) (\sin(wv\Delta^{1-1/\beta}x) - wv\Delta^{1-1/\beta}x1_{\{|x|\leq 1\}}) & \text{if } \beta \geq 1. \end{cases}$$

So we can inject these formulas directly into (79).

As for the asymptotic variance in Theorem 3, it is even simpler. Indeed, we have here

$$(83) \quad \Psi_k(u, 0) = e^{-w^\beta u^\beta / 2}.$$

Therefore $I(k) = -\Psi'_k(1, 0) = \beta w^\beta e^{-w^\beta / 2} / 2 > 0$ and $J(k) = \frac{1}{2}(\Psi_k(2, 0) + 1) - \psi_k(1, 0)^2 = \frac{1}{2}(1 + e^{-(2w)^\beta / 2}) - e^{-w^\beta}$, and thus

$$(84) \quad \Sigma^2(k) = 2 \frac{1 + e^{-(2w)^\beta / 2} - 2e^{-w^\beta}}{\beta^2 w^{2\beta} e^{-w^\beta}}.$$

When $\beta < 2$, it turns out that the minimal variance is achieved for some value $w = w_\beta \in (0, \infty)$, whereas $\Sigma^2(k)$ tends to ∞ when w goes either to 0 or to ∞ . In contrast, when $\beta = 2$ the variance $\Sigma^2(k)$ goes to $1/2$ as $w \rightarrow 0$: recall once more that $1/2$ is the efficient variance in that case.

For the semiparametric situation, things are even simpler. The estimating function of (66) becomes

$$(85) \quad U_{n,G,\beta,k}(u) = V_n(w) - \Psi_k\left(\frac{u}{S_n}, 0\right),$$

provided in (78) we sum over $i \in \{q_n + m_n + 1, \dots, n\}$. Moreover $u \mapsto \Psi_k(u, 0)$ is invertible, so the estimator $\hat{\sigma}_n(k)$ takes the simple explicit form

$$(86) \quad \hat{\sigma}_n(k) = S_n \frac{2^{1/\beta}}{w} \left(-\log \left(\frac{1}{p_n} \sum_{i=q_n+m_n+1}^n \cos \left(\frac{w \Delta_n^{-1/\beta} (\chi_i^n - B_n)}{S_n} \right) \right) \right)^{1/\beta}$$

if the argument of the logarithm is positive (otherwise, put for example $\hat{\sigma}_n(k) = 1$).

7. Example: Power and truncated power functions. Another natural choice for the function k is a power function, that is $k(x) = |x|^r$, for some $r > 0$ when $\beta = 2$ and $r \in (0, \beta/2)$ otherwise (when $\beta = 2$ this is – in principle – optimal for $r = 2$). In general, the function $\Psi_{G,\Delta_n,\beta,k_n}$ is not explicit but can be numerically approximated via Monte–Carlo procedures for example. We can also compute the

limiting variance: with the notation $m_r = E(|W_1|^r)$ we get $I(k) = -rm_r$ and $J(k) = m_{2r} - m_r^2$, hence

$$(87) \quad \Sigma^2(k) = \frac{m_{2r} - m_r^2}{r^2 m_r^2}.$$

When $\beta = 2$ we have a closed expression for m_r (see (95) below), and not surprisingly $\Sigma^2(k)$ achieves its minimum, equal to $1/2$, at $r = 2$: recall that $1/2$ is the “efficient” variance in that case. When $\beta < 2$ we have no explicit expression for these moments. However, $\Sigma^2(k)$ goes to ∞ when r increases to $\beta/2$, and we conjecture that $\Sigma^2(k)$ is monotone increasing in r (this property holds at least when $\beta = 1$); so one should take r as small as possible, although $r = 0$ is of course excluded.

In the semiparametric setting, the previous choice is not admissible, since k has to be bounded. So we must “truncate” the argument, by using the following function $k = k_\gamma$:

$$(88) \quad k_\gamma(x) = |x|^r 1_{\{|x| \leq \gamma\}}$$

for some constant γ . The function $\Psi_{k_\gamma}(u, 0) = u^r E(|W_1|^r 1_{\{|W_1| \leq \gamma/u\}})$ is invertible from a neighborhood I of $u = 1$ onto some interval I' , and we write $\Psi_{k_\gamma}^{-1}(v)$ for the inverse function at $v \in I'$. Then if B_n and S_n are the preliminary estimators, and if

$$(89) \quad V_n(\gamma) = \frac{1}{p_n \Delta_n^{r/\beta}} \sum_{i=m_n+1}^n |\chi_i^n - B_n|^r 1_{\{|\chi_i^n| \leq \gamma \Delta_n^{1/\beta}\}},$$

the estimator $\hat{\sigma}_n(k_\gamma)$ is defined by

$$(90) \quad \hat{\sigma}_n(k_\gamma) = S_n \Psi_{k_\gamma}^{-1} \left(\frac{V_n(\gamma S_n)}{S_n^r} \right)$$

if the argument of $\Psi_{k_\gamma}^{-1}$ above is in I' , and $\hat{\sigma}_n(k_\gamma) = 1$ (for example) otherwise. This is almost as explicit as (86) is. Since k_γ is even we again have $J(k_\gamma) = 0$, whereas

$$(91) \quad \Sigma^2(k_\gamma) = \frac{M_{\gamma,2r} - M_{\gamma,r}^2}{(rM_{\gamma,r} - 2h_\beta(\gamma)\gamma^{r+1})^2}, \quad \text{where } M_{\gamma,s} = \mathbf{E}(|W_1|^s 1_{\{|W_1| \leq \gamma\}}).$$

We can then try to minimize this variance, by appropriately choosing the two constants $\gamma > 0$ and $r > 0$.

One could also use k_{γ_n} , the r th power truncated at some level $\gamma_n > 0$ depending on n : our general results do not apply, but similar results, with possibly other rates, should obviously apply. In fact, in the next section we work out completely this kind of truncated power functions in a particular case, to check that it is best (for the rate of convergence at least) to take a constant level $\gamma_n = \gamma$, as it is implicitly proposed in the method previously developed.

8. Example: Brownian motion plus Gaussian compound Poisson process. In this section, we present a fully worked out example, where W is Brownian motion and Y is a compound Poisson process with Gaussian jumps, say $N(0, \eta)$, and intensity of jumps given by some $\lambda > 0$. [1] and [8] studied the estimation of the parameters of this model, using a variety of methods.

As usual, we are interested in estimating the parameter σ given the increments χ_i^n of $X_t = \sigma W_t + Y_t$ (see (3)). We consider a number of estimating equations for this model, based on the power or truncated power variations

$$(92) \quad V_n(c, \kappa) = \frac{1}{p_n \Delta_n^{r/2}} \sum_{i=m_n+1}^n |\chi_i^n|^r \mathbf{1}_{\{|\chi_i^n| \leq \tau(\Delta_n)\}},$$

for $r \in (0, 2]$. Here $\tau(\Delta)$ is the truncation rate, taken to be of the form $\tau(\Delta) = c\Delta^{1/2+\kappa}$ with c a constant and $\kappa \in (-1/2, \infty)$.

Note that V_n above is exactly $V_n(\gamma)$ of (89) with $\gamma = \Delta_n^{-1/2}\tau(\Delta_n)$ (here Y is symmetrical, so $B_n = 0$). The associated estimator is then given by

$$(93) \quad \hat{\sigma}_n = S_n H_{\Delta_n}^{-1} \left(\frac{V_n(cS_n, \kappa)}{S_n^r} \right)$$

where H_{Δ}^{-1} is the local inverse around 1 of the function $H_{\Delta}(u) = E(|uW_{\Delta}|^r \mathbf{1}_{\{|uW_{\Delta}| \leq \tau(\Delta)\}})$.

When $c = \infty$ we get the (non truncated) r th power variation. If $c < \infty$ and $\kappa = 0$ this corresponds to taking $k = k_c$, as given by (88): we essentially eliminate from the sum above the increments in which Y jumps. When $\kappa > 0$ we eliminate more increments, and fewer when $\kappa < 0$.

The expected values of the powers without truncation are given by

$$(94) \quad \mathbf{E}(|X_\Delta|^r) = \sum_{j=0}^{+\infty} \frac{2^{r/2}}{\sqrt{\pi} j!} \Gamma\left(\frac{1+r}{2}\right) e^{-\lambda\Delta} (\lambda\Delta)^j (\sigma^2\Delta + j\eta)^{r/2},$$

$$(95) \quad \mathbf{E}(|\sigma W_\Delta|^r) = \frac{2^{r/2}}{\sqrt{\pi}} \Gamma\left(\frac{1+r}{2}\right) \sigma^r \Delta^{r/2}$$

With truncation at rate $\tau(\Delta)$, we get

$$(96)$$

$$\begin{aligned} & \mathbf{E}(|X_\Delta|^r 1_{\{|X_\Delta| \leq \tau(\Delta)\}}) \\ &= e^{-\lambda\Delta} \sum_{j=0}^{+\infty} \frac{2^{r/2}}{\sqrt{\pi} j!} \left(\Gamma\left(\frac{1+r}{2}\right) - \Gamma\left(\frac{1+r}{2}, \frac{\tau(\Delta)^2}{2(\sigma^2\Delta + j\eta)}\right) \right) (\lambda\Delta)^j (\sigma^2\Delta + j\eta)^{r/2} \end{aligned}$$

where $\Gamma(a, \cdot)$ denotes the incomplete Gamma function of order a , and

$$(97) \quad \mathbf{E}(|\sigma W_\Delta|^r 1_{\{|\sigma W_\Delta| \leq \tau(\Delta)\}}) = \frac{2^{r/2}}{\sqrt{\pi}} \left(\Gamma\left(\frac{1+r}{2}\right) - \Gamma\left(\frac{1+r}{2}, \frac{\tau(\Delta)^2}{2\sigma^2\Delta}\right) \right) \sigma^r \Delta^{r/2}.$$

When $r = 2$, we have $\Gamma(3/2) = \sqrt{\pi}/2$ and $\Gamma(\frac{3}{2}, x) = e^{-x}\sqrt{x} + \sqrt{\pi}\Phi(\sqrt{2x})$ where Φ denotes the cdf of the $N(0, 1)$ law. Similarly simpler expressions are also obtained in the case where $r = 1$, since $\Gamma(1) = 1$ and $\Gamma(1, x) = e^{-x}$.

As described above, in the semiparametric case where the distribution of Y is not known to the statistician, we propose to use an approximate centering based on computing these expectations assuming that $X = \sigma W$ only (i.e., as if there were no jumps) and we will study the behavior of this estimator when Y is in fact a compound Poisson process. The effect of the misspecification error is to bias the resulting estimator of σ . But, at the leading order in Δ , the expected values of the moments functions computed without jumps coincide with those computed under the correct specification. Indeed, for X from (1), we have

$$\mathbf{E}(|X_\Delta|^r) = \mathbf{E}(|\sigma W_\Delta|^r) + o(\Delta^{r/2})$$

$$\mathbf{E}(|X_\Delta|^r 1_{\{|X_\Delta| \leq \tau(\Delta)\}}) = \mathbf{E}(|\sigma W_\Delta|^r 1_{\{|\sigma W_\Delta| \leq \tau(\Delta)\}}) + o(\Delta^{r/2}),$$

with the second result following from

$$(98) \quad \Gamma(a, x) = \begin{cases} \Gamma(a) + x^a \left(-\frac{1}{a} + \frac{x}{1+a} + O(x^2) \right) & \text{near } 0 \\ e^{-x} x^{-1+a} \left(1 + \frac{a-1}{x} + O(x^{-2}) \right) & \text{near } +\infty. \end{cases}$$

As a result, the bias of the estimator of σ based on approximate centering will vanish asymptotically in Δ and we will have a result of the form

$$\sqrt{n\Delta_n^{v_1}} (\hat{\sigma}_n - \bar{\sigma}_n) \rightarrow N(0, v_0)$$

where

$$\bar{\sigma}_n = \sigma + b_0 \Delta_n^{b_1} + o(\Delta_n^{b_1})$$

with $b_1 > 0$. (If $b_1 = 0$ for some choice of (r, κ, c) then the parameter σ is not identified by an estimating function based on that combination.) Also, $v_1 = 0$ corresponds to a rate of convergence of the estimator of $n^{1/2}$, and any value $v_1 > 0$ corresponds to a slower than $n^{1/2}$ rate of convergence.

We also note that when $b_1 > 0$ the rate of convergence and asymptotic variance of the semiparametric estimator of σ are identical at the leading order in Δ_n to the expressions one would obtain in the fully parametric, correctly specified, case where centering of the estimating equation is done with either (94) or (96) as appropriate, instead of the approximate centering using (95) or (97). Centering using the latter is of course the only feasible estimator in the semiparametric case where the distribution of Y is unknown.

In what follows, we use the explicitness of this model to fully characterize the asymptotic distribution of the semiparametric estimator of σ , i.e., (b_0, b_1, v_0, v_1) as functions of (r, κ, c) and the parameters of the model (σ, λ, η) .

8.1. *Power variations without truncation.* In that situation, we have for the asymptotic variance:

- When $0 < r < 1$, we have $v_1 = 0$ and $v_0 = \frac{1}{r^2} \left(\sqrt{\pi} \frac{\Gamma(\frac{1+r}{2})}{\Gamma(\frac{1+r}{2})^2} - 1 \right)$.
- When $r = 1$, we have $v_1 = 0$ and $v_0 = \frac{1}{2} ((\pi - 2)\sigma^2 + \pi\lambda\eta)$.
- When $1 < r < 2$, we have $v_1 = r - 1$ and $v_0 = \frac{\sqrt{\pi}\sigma^{2-2r}\lambda\eta^r}{r^2} \frac{\Gamma(\frac{1+r}{2})}{\Gamma(\frac{1+r}{2})^2}$.

As for the bias, when $0 < r < 2$ we have $b_1 = 1 - r/2$ and $b_0 = \frac{\sigma^{1-r}\lambda\eta^{r/2}}{r}$.

Remark 11. *The estimator based on power variations converges (not taking the bias into consideration) at rate $n^{1/2}$ only when $r \leq 1$. When $r > 1$ the mixture of jumps and volatility slows down the rate of convergence ($v_1 > 0$). When $r = 2$, the parameter σ is simply not identified, as is obvious from the fact that $E(X_\Delta^2) = (\sigma^2 + \lambda\eta)\Delta$. This is also apparent here from the fact that $b_1 \downarrow 0$ as $r \uparrow 2$, so the bias no longer vanishes asymptotically. And the bias even worsens the rate, of course.*

Remark 12. *When $r < 1$, the asymptotic variance v_0 is identical to the expression obtained without jumps, as was the case when the log-likelihood score was used as an estimating equation. When $r = 1$, the rate of convergence remains $n^{1/2}$, but v_0 is larger in the presence of jumps.*

8.2. *Power variations with $\Delta^{1/2}$ truncation.* If we truncate the increments according to $\tau(\Delta) = c\Delta^{1/2}$, then $v_1 = 0$ for all values of $r \in (0, 2]$ and

$$v_0 = \frac{2^r \sigma^{4+2r} \left(\sqrt{\pi} \left(\Gamma\left(\frac{1}{2} + r\right) - \Gamma\left(\frac{1}{2} + r, \frac{c^2}{2\sigma^2}\right) \right) - \left(\Gamma\left(\frac{1+r}{2}\right) - \Gamma\left(\frac{1+r}{2}, \frac{c^2}{2\sigma^2}\right) \right)^2 \right)}{\left(\sqrt{2}c^{1+r} \exp\left(-\frac{c^2}{2\sigma^2}\right) - 2^{r/2}r\sigma^{1+r} \left(\Gamma\left(\frac{1+r}{2}\right) - \Gamma\left(\frac{1+r}{2}, \frac{c^2}{2\sigma^2}\right) \right) \right)^2}$$

As for the bias, we have $b_1 = 1$ and

$$b_0 = \frac{\sigma\lambda \left(\Gamma\left(\frac{1+r}{2}\right) - \Gamma\left(\frac{1+r}{2}, \frac{c^2}{2\sigma^2}\right) \right)}{\left(\Gamma\left(\frac{1+r}{2}\right) - \Gamma\left(\frac{1+r}{2}, \frac{c^2}{2\sigma^2}\right) \right) - 2 \left(\Gamma\left(\frac{3+r}{2}\right) - \Gamma\left(\frac{3+r}{2}, \frac{c^2}{2\sigma^2}\right) \right)}.$$

Remark 13. *Truncating at rate $\Delta^{1/2}$ restores the convergence rate $n^{1/2}$ for all values of r , (again, regardless of the bias) and permits identification when $r = 2$.*

When $0 < r < 1$ (where the rate $n^{1/2}$ was already achieved without truncation), not truncating can lead to either a smaller or larger value of v_0 than truncating at rate $n^{1/2}$, depending upon the values of (σ^2, c) .

Remark 14. *The asymptotic variance v_0 is identical to its expression when no jumps are present, as it should be in view of our general results (as said before, this type of truncation leads to the estimators studied in our general results). In all cases, the bias is smaller than when no truncation is applied.*

8.3. *Power variations with slower than $\Delta^{1/2}$ truncation.* If we now keep too many increments by truncating according to $\tau(\Delta) = c\Delta^{1/2+\kappa}$, with $-1/2 < \kappa < 0$, then we have for $r \in (0, 2]$:

- When $-3/(2 + 4r) < \kappa < 0$, we have $v_1 = 0$ and

$$v_0 = \frac{\sigma^2}{r^2} \left(\sqrt{\pi} \frac{\Gamma(\frac{1}{2} + r)}{\Gamma(\frac{1+r}{2})^2} - 1 \right)$$

- When $\kappa = -3/(2 + 4r)$, we have $v_1 = 0$ and

$$v_0 = \frac{2^{1/2-r} c^{1+2r} \sqrt{\pi} \lambda \sigma^{2-2r}}{r^2 (1 + 2r) \eta^{1/2} \Gamma(\frac{1+r}{2})^2} + \frac{\sigma^2}{r^2} \left(\sqrt{\pi} \frac{\Gamma(\frac{1}{2} + r)}{\Gamma(\frac{1+r}{2})^2} - 1 \right)$$

- When $-1/2 < \kappa < -3/(2 + 4r)$, we have $v_1 = -\kappa - 2r\kappa - 3/2 > 0$ and

$$v_0 = \frac{2^{1/2-r} c^{1+2r} \sqrt{\pi} \lambda \sigma^{2-2r}}{r^2 (1 + 2r) \eta^{1/2} \Gamma(\frac{1+r}{2})^2}.$$

As for the bias, we have:

- When $-1/(2 + 2r) < \kappa < 0$, we have $b_1 = 1$ and $b_0 = -\frac{\lambda\sigma}{r}$
- When $\kappa = -1/(2 + 2r)$, we have $b_1 = 1$ and $b_0 = \frac{\lambda\sigma}{(1+r)} \left(\frac{2^{1/2-r/2} c^{1+r}}{r\sqrt{\eta}\sigma^r \Gamma(\frac{1+r}{2})} - 1 - \frac{1}{r} \right)$
- When $-1/2 < \kappa < -1/(2 + 2r)$, we have $b_1 = 3/2 + \kappa + r\kappa > 0$ and $b_0 = \frac{2^{1/2-r/2} c^{1+r} \lambda \sigma^{1-r}}{r(1+r)\sqrt{\eta}\Gamma(\frac{1+r}{2})}$.

Remark 15. When $0 < r < 1$, we are automatically in the situation where $\kappa > -3/(2+4r)$, and hence keeping more than $O(\Delta_n^{1/2})$ increments results in the convergence rate $n^{1/2}$ and the same asymptotic variance v_0 as when keeping all increments (i.e., not truncating at all). When $1 < r < 2$, however, it is possible to restore the convergence rate $n^{1/2}$ (compared to not truncating) by keeping more than $O(\Delta_n^{1/2})$ increments, but still “not too many” of them ($-3/(2+4r) \leq \kappa < 0$) beyond that; but even keeping a larger fraction of the increments ($-1/2 < \kappa < -3/(2+4r)$) results in an improvement over keeping all increments since $3/2 - \kappa - 2r\kappa < r - 1$ so that the rate of convergence of $\hat{\sigma}_n$, although slower than $n^{1/2}$, is nonetheless faster than $n^{1/2} \Delta_n^{(r-1)/2}$.

Remark 16. The expressions for $\kappa < 0$ do not converge to those with $O(\Delta_n^{1/2})$ truncation as $\kappa \uparrow 0$ because of the essential singularity of the incomplete Γ function near infinity, given in (98): when $\tau(\Delta) = c\Delta^{1/2+\kappa}$ then $\Gamma((1+r)/2, \cdot)$ is evaluated at $\tau(\Delta)^2/(2\sigma^2\Delta) = c^2\Delta^{2\kappa}/(2\sigma^2)$ and for fixed $\kappa < 0$, terms proportional to $\exp(-c^2\Delta^{2\kappa}/(2\sigma^2))$ are negligible in the Taylor series in Δ of v_0 and b_0 . This is not the case when $\kappa = 0$ however.

Remark 17. As for the bias, keeping “too many” but not all increments ($-1/2 < \kappa < -1/(2+2r)$) leads to a smaller bias than keeping all increments, since $3/2 + \kappa + r\kappa > 1 - r/2$, but to a larger bias than keeping just the right amount since $3/2 + \kappa + r\kappa < 1$.

8.4. *Power variations with faster than $\Delta^{1/2}$ truncation.* Finally, if we keep too few increments by truncating according to $c\Delta^{1/2+\kappa}$, with $\kappa > 0$, then $v_1 = \kappa$ for all values of $r \in (0, 2]$ and

$$v_0 = \frac{\sqrt{2\pi}(1+r)^2\sigma^3}{2c(1+2r)}$$

As for the bias, we have $b_1 = 1$ and $b_0 = \sigma\lambda$.

Remark 18. Truncating at a rate faster than $\Delta^{1/2}$ deteriorates the convergence rate of the estimator from $n^{1/2}$ to $n^{1/2}\Delta_n^{\kappa/2}$: while we successfully eliminate the impact of

jumps on the estimator, we are at the same time reducing the effective sample size utilized to compute the estimator, which increases its asymptotic variance.

Remark 19. *The expressions for v_0 and b_0 for $\kappa > 0$ also do not converge to those with $O(\Delta_n^{1/2})$ truncation as $\kappa \downarrow 0$ because once again we cannot interchange the order of the limits $\Delta_n \rightarrow 0$ and $\kappa \rightarrow 0$.*

8.5. *Comparison with the general case.* Let us compare, in the semiparametric case, the specific results just obtained with the general results obtained in Theorems 4 and 5. In the present situation we have $G \in \overline{\mathcal{G}}'_0$. So these general results assert that if

$$(99) \quad n\Delta_n^2 \rightarrow 0,$$

then the estimators $\hat{\sigma}_n$ converge at a rate \sqrt{n} , and the limit of the normalized error is Gaussian without bias; when (99) fails but $\Delta_n \rightarrow 0$ yet, then the sequence $((\sqrt{n} \wedge \Delta_n^{-1})(\hat{\sigma}_n - \sigma))$ is tight.

The estimators (93) converge at rate \sqrt{n} when $v_1 = 0$ and $n\Delta_n^{2b_1}$ is bounded (then there is a bias) or $n\Delta_n^{2b_1} \rightarrow 0$ (there is no bias). Otherwise, the sequence $(\sqrt{n\Delta_n^{v_1}} \wedge \Delta_n^{-b_1})(\hat{\sigma}_n - \sigma)$ is tight. Then:

- Power variation without truncation: we have a rate \sqrt{n} only when $r \in (0, 1]$ and $n\Delta_n^{2-r}$ is bounded. Otherwise the rate is always worse than in our general results: this was expected, of course.
- Power variation with $\Delta^{1/2}$ truncation: If $n\Delta_n^2 \rightarrow 0$ we have rate \sqrt{n} with asymptotically unbiased error. If $n\Delta_n^2 \rightarrow a \in (0, \infty)$ we have rate \sqrt{n} with asymptotically biased error. If $n\Delta_n^2 \rightarrow \infty$, then $\Delta_n^{-1}(\hat{\sigma}_n - \sigma)$ converges in probability to the constant b_0 : this is a bit better than what we get by applying the general results recalled above. This holds irrespectively of $r \in (0, 2]$ (and

also for $r > 2$ here, as a matter of fact), but of course the asymptotic variance depends on r , and also on c .

- Power variation with slower than $\Delta^{1/2}$ truncation: The rate is \sqrt{n} if $-1/(2 + 2r) \leq \kappa < 0$ and $n\Delta_n^2$ is bounded, or if $-3(2 + 4r) \leq \kappa < -1/(2 + 2r)$ and $n\Delta_n^{3+2\kappa+2r\kappa}$ is bounded. This is worse than the previous case.
- Power variation with faster than $\Delta^{1/2}$ truncation: The rate is at most $\sqrt{n\Delta_n^\kappa}$, and always worst than in the $\Delta^{1/2}$ truncation case.

9. Example: Sum of two stable processes. In this last section we consider the case where Y is also a symmetric stable process, with index $\alpha \in (0, \beta)$. Then $G \in \overline{\mathcal{G}}'_\alpha$.

9.1. *The empirical characteristic function.* First, we can consider estimators based on the empirical characteristic function, that is we consider $k(x) = \cos(wx)$ for some $w > 0$. We have the parametric estimate $\hat{\sigma}_n = \hat{\sigma}_n(G, \phi, k)$ of Theorem 3 (here k is bounded, so ϕ is indeed irrelevant). The sequence $\sqrt{n}(\hat{\sigma}_n - \sigma)$ converges in law to $N(0, \sigma^2 \Sigma^2(k))$, where $\Sigma^2(k)$ is given by (84). On the other hand we have the semiparametric estimators $\hat{\sigma}_n(k)$, which by Theorem 5 behaves as such: under

$$(100) \quad n\Delta_n^{\frac{2(\beta-\alpha)}{\beta}} \rightarrow 0,$$

$\sqrt{n}(\hat{\sigma}_n(k) - \sigma)$ converges in law to $N(0, \sigma^2 \Sigma^2(k))$. And in general the sequence $(\sqrt{n} \wedge \Delta_n^{-\frac{\beta-\alpha}{\beta}})(\hat{\sigma}_n - \sigma)$ is tight.

In fact, since we are in Case 2 the preliminary estimator $S_n = S_n(G)$ is the same in both cases, and $\hat{\sigma}_n$ and $\hat{\sigma}_n(k)$ are the solution of $U_n(u) = 0$ and $U'_n(u) = 0$ respectively, which are closest to S_n , and the difference between these two estimating functions is

$$U_n(u) - U'_n(u) = \hat{U}_n(u) := \Psi_{G, \Delta_n, \beta, k} \left(\frac{u}{S_n}, \frac{1}{S_n}, 0 \right) - \Psi_k \left(\frac{u}{S_n}, 0 \right)$$

(recall (79) and (85)). If we use the explicit forms (80) and (83), we get

$$\widehat{U}_n(u) = e^{-w^\beta u^\beta / 2S_n^b} \left(e^{w^\alpha \Delta_n^{\frac{\beta-\alpha}{\beta}} / 2S_n^\alpha} - 1 \right),$$

which is equivalent to $\frac{w^\alpha}{2\sigma^\alpha} \Delta_n^{\frac{\beta-\alpha}{\beta}} e^{-w^\beta/2}$ as $n \rightarrow \infty$ and $u \rightarrow \sigma$ (recall that $S_n \rightarrow \sigma$ in probability). Since $\Psi'_k(1, 0) = -\beta e^{-w^\beta/2} \neq 0$, we deduce that the difference $\widehat{\sigma}_n(k) - \widehat{\sigma}_n$ is equivalent (in probability) to $-(w^a/2\beta\sigma^\alpha) \Delta_n^{\frac{\beta-\alpha}{\beta}}$. Therefore, in addition to the fact that $\sqrt{n}(\widehat{\sigma}_n(k) - \sigma)$ converges in law to $N(0, \sigma^2 \Sigma^2(k))$ under (100), we get

- If $n \Delta_n^{\frac{\beta-\alpha}{\beta}} \rightarrow a^2 \in (0, \infty)$, then $\sqrt{n}(\widehat{\sigma}_n(k) - \sigma)$ converges in law to $N(-aw^a/2\beta\sigma^\alpha, \sigma^2 \Sigma^2(k))$,
- If $n \Delta_n^{\frac{\beta-\alpha}{\beta}} \rightarrow \infty$, then $\Delta_n^{-\frac{\beta-\alpha}{\beta}}(\widehat{\sigma}_n(k) - \sigma)$ converges in probability to the constant $-w^a/2\beta\sigma^\alpha$.

We conclude that the results of Theorem 5 are sharp, for the particular estimation functions $k(x) = \cos(wx)$ at least.

9.2. Truncated power functions. We can do a similar analysis for the estimators (90), based on the truncated power variation $V_n(\gamma)$ of (89) with $B_n = 0$ (because Y is symmetrical here). That is, we consider the truncated power variations at the level $\Delta_n^{1/\beta}$. Namely when $n \Delta_n^{2\frac{\beta-\alpha}{\beta}} \rightarrow \infty$, one can show that, at least when γ is small enough (but it is probably true for all $\gamma > 0$), then the sequence $\Delta_n^{-\frac{\beta-\alpha}{\beta}}(\widehat{\sigma}_n - \sigma)$ is tight and its limiting distributions include some Dirac masses at non vanishing constants. So here again the results of Theorem 5 are sharp. But of course, as already said before, this does not completely rule out the existence of estimators constructed in a different way and behaving better.

10. Conclusions. We exhibited a class of estimators for the volatility parameter σ in a model where the driving process W_t is perturbed by another process Y_t . These estimators can be designed in such a way that they are immune to the presence of the perturbation Y_t : they are asymptotically efficient, in the strong sense that they behave asymptotically like the efficient estimators for the model $X_t = \sigma W_t$ with no perturbing term Y_t .

REFERENCES

- AÏT-SAHALIA, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics* **74** 487–528.
- AÏT-SAHALIA, Y. and JACOD, J. (2004). Fisher’s information for discretely sampled Lévy processes. Tech. rep., Princeton University and Université de Paris VI.
- FENECH, A. P. (1976). Asymptotically efficient estimation of location for a symmetric stable law. *Annals of Statistics* **4** 1088–1100.
- FEUERVERGER, A. and MCDUNNOUGH, P. (1981b). On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society - Series B* **43** 20–27.
- GODAMBE, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics* **31** 1208–1211.
- HANSEN, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054.
- HEYDE, C.C. (1997). *Quasi-Likelihood and Its Applications*. Springer-Verlag, New York.
- MANCINI, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell’Istituto Italiano degli Attuari* **LXIV** 19–47.
- PRESS, S. J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association* **67** 842–846.
- SINGLETON, K. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* **102** 111–141.
- ZOLOTAREV, V. M. (1986). *One-dimensional Stable Distributions*, vol. 65 of *Translations of Mathematical Monographs*. American Mathematical Society.

DEPARTMENT OF ECONOMICS
PRINCETON UNIVERSITY AND NBER
PRINCETON, NJ 08544-1021
E-MAIL: YACINE@PRINCETON.EDU

LABORATOIRE DE PROBABILITÉS (UMR 7599)
UNIVERSITÉ P. ET M. CURIE (PARIS-6)
75252 PARIS CÉDEX 05
E-MAIL: JJ@CCR.JUSSIEU.FR