The Annals of Statistics
2007, Vol. 35, No. 1, 355–392
DOI: 10.1214/009053606000001190

VOLATILITY ESTIMATORS FOR DISCRETELY SAMPLED LÉVY PROCESSES

BY YACINE AÏT-SAHALIA[1] AND JEAN JACOD

Princeton University and Université de Paris-6

This paper studies the estimation of the volatility parameter in a model where the driving process is a Brownian motion or a more general symmetric stable process that is perturbed by another Lévy process. We distinguish between a parametric case, where the law of the perturbing process is known, and a semiparametric case, where it is not. In the parametric case, we construct estimators which are asymptotically efficient. In the semiparametric case, we can obtain asymptotically efficient estimators by sampling at a sufficiently high frequency, and these estimators are efficient uniformly in the law of the perturbing process.

1. Introduction. Models allowing for sample path discontinuities or jumps are becoming increasingly popular, especially in mathematical finance. Among jump processes, Lévy processes play a central role due to their analytical tractability and their ability to span the behavior of most discontinuous processes. However, even for Lévy processes, relatively little is known about the corresponding inference problem, especially for high frequency data. Specifically, suppose that a Lévy process X, say the log-price of a financial asset, is observed at n times $\Delta_n, 2\Delta_n, \dots, n\Delta_n$. Since $X_0 = 0$, this amounts to observing the n increments $\chi_i^n = X_{i\Delta_n} - X_{(i-1)\Delta_n}$. Their density, and hence the corresponding likelihood function and Fisher information, are not known in closed form. Moreover, under the natural asymptotics for high frequency data, where the sampling interval $\Delta_n \to 0$, these densities explode.

When $\Delta > 0$ is fixed, we are on familiar ground. We observe n i.i.d. variables distributed as X_Δ. If, further, this variable has a density which depends smoothly on the parameter of interest, η, the Fisher information at stage n has the form $I_{n,\Delta}(\eta) = nI_\Delta(\eta)$, where $I_\Delta(\eta) > 0$ is the Fisher information of the model based on the observation of the single variable X_Δ, we have the LAN property with rate $\sqrt{n}$, the asymptotically efficient estimators $\widehat{\eta}_n$ are those for which $\sqrt{n}(\widehat{\eta}_n - \eta)$ converges in law to the normal distribution $N(0, I_\Delta(\eta)^{-1})$ and the MLE solves the problem (see, e.g., [5–7]). In this setting, a variety of other methods have been proposed in the literature: using the empirical characteristic function as an estimating

Received January 2005; revised March 2006.

[1]Supported in part by NSF Grants SBR-03-50772 and DMS-05-32370.

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

Key words and phrases. Jumps, efficiency, inference, discrete sampling.

equation (see, e.g., [8, 10, 18, 19] and Chapter 4 in [23]), maximum likelihood by Fourier inversion of the characteristic function (see [9]), a regression based on the explicit form of the characteristic function (see [14]) or other numerical approximations (see [16, 17]). Some of these methods were compared in [3].

Things become more complicated when sampling takes place at increasingly higher frequency, that is when $\Delta_n \to 0$. Here, the χ_i^n's are i.i.d. for any given n, but their laws depend on n. The Fisher information at stage n still has the form $I_{n,\Delta_n}(\eta) = nI_{\Delta_n}(\eta)$, but the asymptotic behavior of $I_{\Delta_n}(\eta)$ is far from obvious and estimating all of the parameters of a general Lévy process in this setting still remains out of our reach. So, in this paper we study an example which, despite its apparent simplicity, turns out to be already quite complex. We suppose that

$$X_t = \sigma W_t + Y_t, \tag{1}$$

where $\sigma > 0$ is the parameter of interest, W is a standard symmetric stable process with index $\beta \in (0, 2]$ and Y is another Lévy process, independent of W and viewed as a perturbation of W. We assume that Y is dominated by W in a sense to be stated below: for example, when W is a Wiener process ($\beta = 2$), this just means that Y has no Brownian part; when $\beta < 2$, Y could, for example, be another stable process with index $\alpha < \beta$ or a compound Poisson process. For instance, in many financial applications, W is a Wiener process and Y may represent frictions that are due to the mechanics of the trading process, in which case, Y would have infinite activity. Alternatively, σW could represent the ordinary fluctuations of the asset value and Y the infrequent arrival of information related to the asset, in which case Y is a compound Poisson process. A number of papers are devoted to the estimation of the integrated volatility $\int_0^t \sigma_s^2 ds$ in the model $X_t = \int_0^t \sigma_s dW_s + Y_t$, where W is a Wiener process, Y is typically a compound Poisson or other specific Lévy process and σ may be stochastic (see, e.g., [4, 15, 21, 22]).

We start with the fully parametric case, where the law of Y is given. Viewing Y as a perturbation of W, our interest then lies in deciding whether we can estimate the parameter σ with the same degree of accuracy as when the process Y is absent, at least asymptotically. The answer to this question is "yes," which we show by analyzing the Fisher information, proving the LAN property and exhibiting efficient estimators, in the strong sense that asymptotically they behave as well as when Y is absent. When W is a Wiener process, this means that one can distinguish between the jumps due to Y and the continuous part of X; this fact was already known in some specific examples (see [1]). It comes as more of a surprise when $\beta < 2$—we can then discriminate between the jumps due to W and those due to Y, despite the fact that both processes jump and we have only discrete observations. This surprising property is our main motivation for studying the case when W is more general than a Wiener process (but dominates Y).

But, given the nature of the problem, we would rather not fully specify the law of Y, so we treat it as a nuisance parameter. This gives rise to a semiparametric situation. There, we show that obtaining asymptotically efficient estimators of σ requires Δ_n to convergence sufficiently fast to 0, but we can then exhibit estimators that are uniformly efficient 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.

In both the parametric and semiparametric cases, 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 simpler estimators which achieve the efficient rate of convergence.

In this paper, we focus on a single parameter, σ. In a companion paper, [2], we study the optimal rate at which other parameters of the model, namely β and a scale parameter θ for the Y process, can be estimated. While we show here that σ can be estimated optimally at rate $n^{1/2}$, independently of the specification of the Y process, this is not the case in general for β and θ. In particular, the rate for β is faster than $n^{1/2}$, and is also unaffected by the presence of Y, but the rate for θ is strongly dependent upon the precise nature of the Y process, and is affected by the presence of the W process.

The paper is organized as follows. In Section 2 we outline the problem and define the class of processes Y that are dominated by W. In Section 3 we summarize the statistical properties in the baseline case where $X_t = \sigma W_t$. In Section 4 we exhibit the behavior of the Fisher information and prove the LAN property. We construct our classes of estimators in the parametric and semiparametric cases and state their asymptotic properties in Sections 5 and 6, respectively. In Sections 7, 8, 9 and 10 we study a number of examples in some detail. The proofs are given in the last three sections.

2. Setup. The process W is a standard symmetric stable process with index $\beta \in (0, 2]$. This means that if $\beta = 2$, then W is a Wiener process and if $\beta < 2$, then its characteristic function is $E(e^{iuW_t}) = e^{-t|u|^\beta}$. The Lévy process Y is independent of W and its law is entirely specified by the law G_Δ of the variable Y_Δ for any given $\Delta > 0$, for instance, Y_1. We write $G = G_1$ and recall that the characteristic function of G_Δ is given by the Lévy–Khintchine formula

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

where (b, c, F) is the "characteristic triple" of G (or of Y): $b \in \mathbb{R}$ is the drift, $c \ge 0$ is the local variance of the continuous part of Y and F is the Lévy jump measure of Y, which satisfies $\int (1 \wedge x^2) F(dx) < \infty$ (see, e.g., Chapter II.2 in [12]).

We define the "domination" of Y by W in terms of the property that G belongs to one of the classes defined below for some $\alpha \le \beta$. Let $\overline{\Phi}$ be the class of all nonnegative continuous functions on $[0,1]$ and Φ be the set of all $\phi \in \overline{\Phi}$ with $\phi(0) = 0$. If $\phi \in \overline{\Phi}$, then we set

$$\mathcal{G}(\phi,\alpha) = \text{the set of all infinitely divisible distributions with } c = 0 \text{ and}$$

$$\forall x \in (0,1] \qquad \begin{cases} x^\alpha F([-x,x]^c) \le \phi(x), & \text{if } \alpha < 2, \\ x^2 F([-x,x]^c) \le \phi(x) \text{ and} & \\ \displaystyle\int_{\{|y|\le x\}} |y|^2 F(dy) \le \phi(x), & \text{if } \alpha = 2, \end{cases} \tag{3}$$

$$\mathcal{G}'(\phi,\alpha) = \{G \in \mathcal{G}(\phi,\alpha),\ G \text{ is symmetrical about } 0\}, \tag{4}$$

$$\mathcal{G}_\alpha = \bigcup_{\phi\in\Phi} \mathcal{G}(\phi,\alpha), \qquad \overline{\mathcal{G}}_\alpha = \bigcup_{\phi\in\overline{\Phi}} \mathcal{G}(\phi,\alpha), \qquad \overline{\mathcal{G}}'_\alpha = \bigcup_{\phi\in\overline{\Phi}} \mathcal{G}'(\phi,\alpha). \tag{5}$$

Observe that as $x \downarrow 0$, we always have

$$\int_{\{|y|\le x\}} |y|^2 F(dy) \to 0 \quad \text{and} \quad x^2 F([-x,x]^c) \to 0,$$

hence

$$\begin{aligned} \alpha \in (0,2] \quad &\Longrightarrow \quad \mathcal{G}_\alpha = \Big\{G \text{ is infinitely divisible},\ c = 0, \\ &\qquad\qquad \lim_{x\downarrow 0} x^\alpha F([-x,x]^c) = 0\Big\}, \\ \alpha = 2 \quad &\Longrightarrow \quad \mathcal{G}_2 = \overline{\mathcal{G}}_2 = \{G \text{ is infinitely divisible},\ c = 0\}. \end{aligned} \tag{6}$$

The definition (6) for $\mathcal{G}_\alpha$ certainly appears simpler than (5), but we will need each class $\mathcal{G}(\phi,\alpha)$ separately for the purpose of stating precise uniformity results for the Fisher information and estimators below.

Note that $\alpha < \alpha'$ implies that $\mathcal{G}_\alpha \subset \overline{\mathcal{G}}_\alpha \subset \mathcal{G}_{\alpha'}$. If G is a (not necessarily symmetric) stable law with index γ, then it belongs to $\mathcal{G}_\alpha$ for all $\alpha > \gamma$, but not to $\mathcal{G}_\gamma$. If Y is a compound Poisson process plus a drift, then $G \in \bigcup_{\alpha>0} \mathcal{G}_\alpha$.

We let $P_{\sigma,G}$ denote the law of the process X in (1). We observe the n i.i.d. increments $\chi_i^n = X_{i\Delta_n} - X_{(i-1)\Delta_n}$. These variables have densities which depend smoothly on σ. The Fisher information of our experiment is $I_{n,\Delta_n}(\sigma,G) = nI_{\Delta_n}(\sigma,G)$, where $I_\Delta(\sigma,G)$ is the Fisher information associated with the observation of a single variable X_Δ, which we will compute below.

Let us also recall what the LAN (local asymptotic normality) property means in this context. Denote by $Z_n(\sigma'|\sigma,G)$ the log–likelihood of the law of the sequence $(\chi_i^n : 1 \le i \le n)$ under $P_{\sigma',G}$ relative to its law under $P_{\sigma,G}$. We say that LAN holds at σ with rate $\sqrt{n}$ and asymptotic Fisher information $I > 0$ if there are random variables $U_n = U_n(\sigma)$ converging in law under $P_{\sigma,G}$ to an $N(0,1)$ variable and such that for any real u, we have $Z_n(\sigma + u/\sqrt{n}|\sigma,G) - u\sqrt{I}U_n + u^2 I/2 \to 0$ in $P_{\sigma,G}$-probability. A sequence $\widehat{\sigma}_n$ is then efficient for estimating $\sigma > 0$ if $\sqrt{n}(\widehat{\sigma}_n - \sigma)$ converges in law under $P_{\sigma,G}$ to an $N(0,1/I)$ variable.

3. The baseline case where Y is absent. Because we will show that the estimation of σ in the presence of Y is asymptotically as good as when $Y = 0$, let us start with the situation where we observe the stable process $X = \sigma W$ with scale parameter $\sigma > 0$ and index parameter $\beta \in (0, 2]$. The variables $n^{1/\beta}\chi_i^n$ are then i.i.d., with the same law as σW_1. In addition, W_1 has a density h_β which is C^∞ and the nth derivative $h_\beta^{(n)}$ behaves as

$$|h_\beta^{(n)}(w)| \sim \begin{cases} \dfrac{c_\beta(1+\beta)(2+\beta)\dots(n+\beta)}{|w|^{n+1+\beta}}, & \text{if } \beta < 2, \\ |w|^n e^{-w^2/2}/\sqrt{2\pi}, & \text{if } \beta = 2, \end{cases} \qquad \text{as } |w| \to \infty, \tag{7}$$

where c_β is a positive constant. Let us also associate with h_β the functions

$$\breve{h}_\beta(w) = h_\beta(w) + w h_\beta^{(1)}(w), \qquad \widetilde{h}_\beta(w) = \frac{\breve{h}_\beta(w)^2}{h_\beta(w)}, \\ \overline{h}_\beta(w) = \frac{w h_\beta^{(1)}(w)}{h_\beta(w)}. \tag{8}$$

Then $\tilde{h}_\beta$ is positive, even and continuous and $\tilde{h}_\beta(w) = 0(1/|w|^{1+\beta})$ as $|w| \to \infty$. Hence, $\tilde{h}_\beta$ is Lebesgue integrable. Moreover, the variable σW_1 admits the density $h_\beta(x/\sigma)/\sigma$. Then:

- The Fisher information $I_\Delta(\sigma, \delta_0)$ associated with the observation of the single variable X_Δ does not depend on Δ and equals $\mathcal{I}(\beta)/\sigma^2$, where

$$\mathcal{I}(\beta) = \int \widetilde{h}_\beta(w)\, dw, \tag{9}$$

  which is well defined and positive. Moreover, if $\beta = 2$ (W is then Brownian motion), then $h_2^{(1)}(w) = -w h_2(w)$ and $\widetilde{h}_2(w) = (1 - w^2 + w^4)h_2(w)$ and thus $\mathcal{I}(2) = 2$.
- Since $h_\beta(x/\sigma)/\sigma$ is a smooth function of σ, we have the LAN property at any $\sigma > 0$ [for the observation of the increments $(\chi_i^n : 1 \le i \le n)$] with rate $\sqrt{n}$ and asymptotic Fisher information $\mathcal{I}(\beta)/\sigma^2$.
- The MLE $\widehat{\sigma}_n$ is asymptotically efficient, in the sense that $\sqrt{n}(\widehat{\sigma}_n - \sigma)$ converges under P_{σ,δ_0} to an $N(0, \sigma^2/\mathcal{I}(\beta))$ variable. In fact, $\widehat{\sigma}_n$ is a positive solution to the equation $H_n(u) = -1$, where $H_n(u) = n^{-1}\sum_{i=1}^n \overline{h}_\beta(\chi_i^n/u)$, and such a solution always exists because H_n is continuous, tends to 0 when $u \to \infty$ and tends to $-(1+\beta)$ when $u \to 0$.

4. Fisher information and LAN. We now return to the situation where Y is present. The first result is a purely parametric one, when the law G of Y_1 is known.

THEOREM 1. *For any* $\Delta > 0$, *we have*

$$I_\Delta(\sigma, G) \le \frac{1}{\sigma^2}\mathcal{I}(\beta) \tag{10}$$

and when $G \in \mathcal{G}_\beta$, *we have, as* $\Delta \to 0$,

$$I_\Delta(\sigma, G) \to \frac{1}{\sigma^2}\mathcal{I}(\beta). \tag{11}$$

Furthermore, if $G \in \mathcal{G}_\beta$ *and* $\Delta_n \to 0$, *then the LAN property holds at any* $\sigma > 0$, *with rate* $\sqrt{n}$ *and asymptotic Fisher information* $\mathcal{I}(\beta)/\sigma^2$, *for the experiments which consist of observing* $(\chi_i^n : 1 \le i \le n)$.

The second result concerns the uniformity of the previous convergence when G varies, so it is a semiparametric result. We will only state the result for the Fisher information.

THEOREM 2. (a) *For any* $\phi \in \Phi$, $\alpha \in (0, \beta]$ *and* $\varepsilon \in (0, 1)$, *we have, as* $\Delta \to 0$,

$$\sup_{G \in \mathcal{G}(\phi,\alpha), \sigma \in [\varepsilon, 1/\varepsilon]} \left| I_\Delta(\sigma, G) - \frac{\mathcal{I}(\beta)}{\sigma^2} \right| \to 0. \tag{12}$$

(b) *For each* n, *let* G^n *denote the standard symmetric stable law of index* α_n, *with* α_n *a sequence strictly increasing to* β. *Then for any sequence* $\Delta_n \to 0$ *such that* $(\beta - \alpha_n)\log \Delta_n \to 0$ (*i.e., the rate at which* $\Delta_n \to 0$ *is sufficiently slow*), *the* $I_{\Delta_n}(\sigma, G^n)$ *converge to a limit strictly less than* $\mathcal{I}(\beta)/\sigma^2$.

In other words, as soon as Y is dominated by W, the presence of Y has no impact on the information terms I_Δ: in the limit where $\Delta \to 0$, the parameter σ can be estimated with exactly the same degree of precision whether Y is present or not. Moreover, part (a) of Theorem 2 states that the convergence of the Fisher information is uniform on the set $\mathcal{G}(\phi, \alpha)$ for all $\alpha \le \beta$. This settles the case where G is considered as a nuisance parameter for the purpose of estimating σ.

But, as soon as α tends to β, we see in part (b) that the convergence disappears. This suggests that the class $\mathcal{G}_\beta$ is effectively the largest one for which the presence of Y does not affect the estimation of σ. For example, if $\beta = 2$, then in part (b), take G^n to be the symmetric stable law with index $\alpha_n \in (0, 2)$ and scale parameter s, in the sense that its characteristic function is $u \mapsto \exp(-s^2|u|^{\alpha_n}/2)$. Then if $\alpha_n \to 2$, for all sequences $\Delta_n \to 0$ satisfying $(2 - \alpha_n)\log \Delta_n \to 0$, we have $I_{\Delta_n}(\sigma, G^n) \to 2/(\sigma^2 + s^2)$. This is, of course, to be expected since in the limit we are observing $\sqrt{\sigma^2 + s^2}W$ and we supposedly know s and wish to estimate σ.

5. Estimation in the parametric case. Here, we suppose that G is known and belongs to $\mathcal{G}_\beta$ and we seek estimators for σ which are efficient in the sense of Theorem 1.

5.1. *Some notation.* We will need to center the increments χ_i^n by accounting for the drift b of Y, as well as the drift coming from the asymmetry of the Lévy measure F, when present. So, when $G \in \mathcal{G}_\alpha$ with $\alpha \leq \beta$, we set

$$b'(G,\alpha) = \begin{cases} b - \int_{\{|x|\leq 1\}} xF(dx), & \text{if } \alpha < 1, \\ b, & \text{if } \alpha \geq 1. \end{cases} \tag{13}$$

Let $Z_\Delta(\alpha) = \Delta^{-1/\beta}(Y_\Delta - b'(G,\alpha)\Delta)$ and let $G'_{\Delta,\alpha}$ denote the law of $Z_\Delta(\alpha)$. Then we define the centered and scaled increments

$$\chi_i'^n(G) = \Delta_n^{-1/\beta}\big(\chi_i^n - b'(G,\beta)\Delta_n\big). \tag{14}$$

Next, for the purpose of constructing estimating equations, if $u > 0$, $v \geq 0$, $z \in \mathbb{R}$ and k is a bounded function, we define

$$\begin{aligned} \Psi_{G,\Delta,\alpha,k}(u,v,z) &= \int h_\beta(x)\,dx \int G'_{\Delta,\alpha}(dw)k(ux+vw+z), \\ \Psi_k(u,z) &= \int h_\beta(x)k(ux+z)\,dx, \end{aligned} \tag{15}$$

so $\Psi_k(u,z) = \Psi_{G,\Delta,\alpha,k}(u,0,z)$ for all G, Δ and α. Then we introduce the "tail function,"

$$\psi(u) = P(|W_1| > 1/u) = 2\int_{1/u}^{\infty} h_\beta(x)\,dx \tag{16}$$

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

Finally, for $\alpha \in (0,2]$ and the function $\phi \in \overline{\Phi}$ defining the class $\mathcal{G}(\phi,\alpha)$, let $\phi'(x) = \sup(\phi(y) : y \in [0,x])$ for $x \in [0,1]$ and

$$\phi_\alpha(x) = \begin{cases} \dfrac{\phi'(x)}{1-\alpha}, & \text{if } \alpha < 1, \\ \phi'(x) + \dfrac{\phi'(x)}{\sqrt{\log(1/x)}} + \phi\big(1 \wedge e^{-\sqrt{\log(1/x)}}\big), & \text{if } \alpha = 1, \\ \phi'(x) + \dfrac{\phi'(\sqrt{x})}{\alpha-1} + \dfrac{\phi'(1)}{\alpha-1}x^{\frac{\alpha-1}{2}}, & \text{if } \alpha > 1 \end{cases} \tag{17}$$

[with $\exp(-\sqrt{\log(1/x)}) = 0$ if $x = 0$]. This obviously defines an increasing function $\phi_\alpha \in \overline{\Phi}$, where $\phi \leq \phi_\alpha$ and also $\phi_\alpha \in \Phi$ whenever $\phi \in \Phi$.

5.2. *Construction of the estimators.* Recall that, for now, $G \in \mathcal{G}_\beta$ is known and so, in particular, we know the centering number $b'(G,\beta)$ used in the centering of the increments (14) and also that $G \in \mathcal{G}(\phi,\beta)$ for some $\phi \in \Phi$. In general, we need to begin with a preliminary estimator $S_n(G)$ for σ which is designed to rule out explosions. If we are willing to assume that σ belongs to the interval $[\varepsilon, 1/\varepsilon]$ for some $\varepsilon \in (0, 1)$, then it would be possible to dispense with the preliminary estimator $S_n(G)$ by defining the estimating function in (24) below slightly differently. Otherwise, we choose an arbitrary sequence m_n of integers satisfying

$$m_n \uparrow \infty, \qquad \frac{m_n}{n} \to 0, \tag{18}$$

and recalling (14) and (16), we define the preliminary estimator as follows:

$$\begin{aligned} V_n(G) &= \frac{1}{m_n} \sum_{i=1}^{m_n} 1_{\{|\chi_i'^n(G)|>1\}}, \\ S_n(G) &= \begin{cases} \psi^{-1}(V_n(G)), & \text{if } 0 < V_n(G) < 1, \\ 1, & \text{otherwise.} \end{cases} \end{aligned} \tag{19}$$

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

$$\sup_x \frac{|k(x)|}{1+|x|^\gamma} < \infty, \qquad I(k) := \int \check{h}_\beta(x) k(x)\, dx \neq 0, \tag{20}$$

where γ satisfies

$$\begin{aligned} &\gamma \in [0, +\infty), \qquad \text{if } \beta = 2, \\ &\gamma \in [0, \beta/2), \qquad \text{if } \beta < 2. \end{aligned} \tag{21}$$

Then we set

$$k_n(x) = \begin{cases} k(x), & \text{if } k \text{ is bounded}, \\ k(x) 1_{\{|k(x)| \le \nu_n\}}, & \text{otherwise}, \end{cases} \tag{22}$$

where ν_n is an increasing sequence of numbers satisfying

$$\nu_n \to \infty, \qquad \nu_n^2 \phi_\beta(\Delta_n^{1/\beta}) \to 0, \qquad \frac{\nu_n^4}{n} \to 0 \tag{23}$$

and where ϕ_β is associated with ϕ [a function such that $G \in \mathcal{G}(\phi, \beta)$] by (17). Conditions (23) limit the growth of k_n and play purely technical roles such as ensuring that the CLT holds.

Then, with $p_n = n - m_n$, and since each k_n is bounded, we can define the estimation functions (for $u > 0$)

$$U_{n,G,\phi,k}(u) = \frac{1}{p_n} \sum_{i=m_n+1}^{n} k_n\left(\frac{\chi_i'^n(G)}{S_n(G)}\right) - \Psi_{G,\Delta_n,\beta,k_n}\left(\frac{u}{S_n(G)}, \frac{1}{S_n(G)}, 0\right). \tag{24}$$

Finally, the estimators for σ are

$$\widehat{\sigma}_n(G, \phi, k) = \begin{cases} \text{the } u > 0 \text{ with } U_{n,G,\phi,k}(u) = 0 \text{ closest to } S_n(G), \\ \qquad \text{if it exists}, \\ 1, \qquad \text{otherwise.} \end{cases} \tag{25}$$

As the notation suggests, this estimator depends on G and k and it depends on ϕ through the choice for k_n made in (23). Of course, it also depends on β, but since β is always fixed, we leave this dependence implicit in the notation.

5.3. *Asymptotic distribution.* With k and $I(k)$ defined as in (20), the following are two finite numbers:

$$J(k) = E(k(W_1)^2) - (E(k(W_1)))^2, \qquad \Sigma^2(k) = \frac{J(k)}{I(k)^2}. \tag{26}$$

THEOREM 3. *Let $\phi \in \Phi$ and let k be a function satisfying* (20) *for some γ satisfying* (21). *Suppose, further, that $\Delta_n \to 0$.*

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

(b) *We have $\Sigma^2(k) \geq 1/\mathcal{I}(\beta)$ and this inequality is an equality if we choose $k = \overline{h}_\beta$, that is, the σ-score from the density of σW_Δ alone* (*without Y*).

REMARK 1. It is, of course, possible (and advisable) to select the function k so as to minimize $\Sigma^2(k)$. The choice $k = \overline{h}_\beta$ is indeed possible: the function $k = \overline{h}_\beta$ satisfies (20) with $\gamma = 0$ (resp. $\gamma = 2$) if $\beta < 2$ (resp. $\beta = 2$). Such a choice gives asymptotically efficient estimators according to the optimality associated with the LAN property (Theorem 1), which, furthermore, 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 to use, we would need to numerically compute the function $\Psi_{G,\Delta,\beta,k}(u, v, 0)$ for a single value $v = 1/S_n(G)$ and all values of u (in principle). Except in special situations, there is no closed form for this function and we would 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. In general, this is not the case for the function $k = \overline{h}_\beta$, except in the important case where $\beta = 2$.

REMARK 3. As an example, we can take $k(x) = |x|^r$ for some $r > 0$ when $\beta = 2$ and some $r \in (0, \beta/2)$ otherwise (when $\beta = 2$ and $r = 2$, this is the optimal choice since $\overline{h}_2(x) = -x^2$): the function $\Psi_{G,\Delta_n,\beta,k_n}$ is still not explicit, but is easily approximated by Monte Carlo techniques, at least when Y_t can be easily simulated. We discuss that choice in some detail in Section 8. Another possibility is to use the empirical characteristic function of the sampled increments, which leads to a closed form expression for $\Psi_{G,\Delta_n,\beta,k_n}$. This will be done in Section 7.

6. Estimation in the semiparametric case. Perhaps more realistic is the case where we seek to estimate σ, but the measure G is unknown. Because of Theorem 2(b), one cannot hope for estimators that are as efficient as in the absence of Y, or even consistent, if all we know is that $G \in \mathcal{G}_\beta$. So, we will assume that G, although unknown, belongs to the class $\overline{\mathcal{G}}_\alpha$ for some $\alpha < \beta$. Since G is unknown, the estimating equations must be based on the law of W alone (and neither on G nor even on α). The challenge is then to achieve rate efficiency, despite the sparse information we have about G. We will exhibit estimators which are efficient if the sampling interval Δ_n converges sufficiently fast to 0. The closer α is to β, the faster Δ_n needs to converge to 0.

6.1. *Construction of the estimators.* We refer to the (simpler) situation where we know that $G \in \overline{\mathcal{G}}'_\alpha$ (that is, $G \in \overline{\mathcal{G}}_\alpha$ and G is symmetrical) and to the situation where we know only that $G \in \overline{\mathcal{G}}_\alpha$ as the *symmetrical* and *asymmetrical* cases, respectively. The construction appears very similar to the one in the parametric case, except that in the asymmetrical case, we now need to produce an estimator B_n for the drift $b'(G, \alpha)$ in order to remove it. Of course, in the symmetrical case, since we know that $b'(G, \alpha) = 0$, we just set $B_n = 0$. In the asymmetrical case, we set $r_n = [\delta n]$ for some arbitrary $\delta \in (0, 1/2)$ ($[x]$ denotes the integer part of x) so that $r_n \sim \delta n$. Then we choose a C^∞, strictly increasing and odd function θ, with bounded derivative, $\theta(0) = 0$ and $\theta(\pm\infty) = \pm 1$ [e.g., $\theta(x) = 2\arctan(x)/\pi$] and set for $u \in \mathbb{R}$,

$$R_n(u) = \frac{1}{r_n}\sum_{i=1}^{r_n} \theta\big(\Delta_n^{-1/\beta}(\chi_i^n - u)\big). \tag{27}$$

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

$$B_n = \inf(u : R_n(u) = 0) \qquad [= \text{the only root of } R_n(\cdot) = 0]. \tag{28}$$

Next, we construct our preliminary estimator for σ. Exactly as in the parametric case, if we are willing to assume that σ lies in a given interval $[\varepsilon, 1/\varepsilon]$, then this preliminary estimator is not needed and can be replaced by $S_n = 1$. In the symmetrical case, with m_n satisfying (18), we set $q_n = 0$ and $p_n = n - m_n$. In the asymmetrical case, with m_n satisfying (18) and r_n as in (27), we set $q_n = r_n$ and $p_n = n - m_n - r_n$. Then in both cases we set

$$\begin{aligned} V_n &= \frac{1}{m_n}\sum_{i=q_n+1}^{q_n+m_n} 1_{\{|\Delta_n^{-1/\beta}(\chi_i^n - B_n)| > 1\}}, \\ S_n &= \begin{cases} \psi^{-1}(V_n), & \text{if } 0 < V_n < 1, \\ 1, & \text{otherwise.} \end{cases} \end{aligned} \tag{29}$$

To construct estimating equations for σ, we choose a function k satisfying (20) with $\gamma = 0$ [that is, k is bounded and $I(k) \neq 0$]. With the notation Ψ_k of (15), we define the estimating functions (for $u > 0$)

$$U_n(u) = \frac{1}{p_n} \sum_{i=q_n+m_n+1}^{n} k\left(\frac{\Delta_n^{-1/\beta}(\chi_i^n - B_n)}{S_n}\right) - \Psi_k\left(\frac{u}{S_n}, 0\right) \tag{30}$$

and the final estimators

$$\widehat{\sigma}_n(k) = \begin{cases} \text{the } u \text{ with } U_n(u) = 0 \text{ which is closest to } S_n, & \text{if it exists,} \\ 1, & \text{otherwise.} \end{cases} \tag{31}$$

Note that unlike the centering utilized for the estimating equation in the parametric case [recall (24)], the centering we now use, based on $\Psi_k(u/S_n, 0)$ in (30), 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. As a result, they are much easier to compute than in the parametric case. This is particularly true when $k(x) = \cos(wx)$ for some $w > 0$ since then $\Psi_k(u, 0) = \exp(-w^\beta u^\beta/2)$ is invertible in u. We will detail this example in the next section, but it is also true in general: first, because the estimators 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 $u = 1$.

6.2. *Asymptotic distribution.* Recall the notation $I(k)$, $J(k)$, $\Sigma^2(k)$ of (20) and (26) and define

$$\rho(\alpha, \beta) = \frac{2(\beta - \alpha)}{\beta(2 + \alpha)}, \qquad \rho'(\alpha, \beta) = \frac{\beta - \alpha}{\beta}. \tag{32}$$

Observe that $\rho(\alpha, \beta) < \rho'(\alpha, \beta)$ when $\alpha < \beta$. The two theorems below cover the properties of the estimators in the symmetrical and asymmetrical cases, respectively.

THEOREM 4. *Let $\alpha \in (0, \beta)$, $\phi \in \overline{\Phi}$, k be a bounded function with $I(k) \neq 0$ and $\varepsilon \in (0, 1)$. Take the symmetrical version of the estimators.*

(a) *If*

$$\sup_n n\Delta_n^{2\rho'(\alpha,\beta)} \to 0, \tag{33}$$

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

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

THEOREM 5. *Let* $\alpha \in (0, \beta)$, $\phi \in \overline{\Phi}$, k *be a bounded function with* $I(k) \neq 0$ *and* $\varepsilon \in (0, 1)$. *Take the asymmetrical version of the estimators.*

(a) *If*

$$\sup_n n \Delta_n^{2\rho(\alpha,\beta)} \to 0, \tag{34}$$

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

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

The optimal choice of the function k was discussed after Theorem 3: when $\beta < 2$, we still have asymptotic efficiency in the situation of Theorem 4(a), provided that $k = \overline{h}_\beta$. When $\beta = 2$, the choice $k = \overline{h}_\beta$ is not permitted here, but with truncation as in $k(x) = -x^2 1_{\{|x| \leq A\}}$, one achieves an asymptotic variance which approaches the optimal variance when A goes to infinity; see Section 8.

REMARK 4. When α increases, then $\rho(\alpha, \beta)$ and $\rho'(\alpha, \beta)$ decrease, so (33) and (34) are more difficult to obtain and the rate in (b) of the two theorems above gets worse, as it should. Also, when (34) fails, the actual rate of convergence [that is, a sequence δ_n such that the law of $\delta_n(\widehat{\sigma}_n(k) - \sigma)$ converges to a nondegenerate limit, or at least admits some nondegenerate weak limiting measures] is not only unknown, but actually depends on the true underlying measure G and, in particular, on the minimal index α' such that $G \in \overline{\mathcal{G}}_{\alpha'}$ (we know $\alpha' \leq \alpha$, but the inequality could be strict). In other words, the rate could be, for example, $\sqrt{n}$ for a particular G, even without (34).

REMARK 5. However, we will see in the examples below (see Section 10 in particular) that (33) is necessary for convergence to a centered distribution with rate $\sqrt{n}$ and also that the rate in (b) of Theorem 4 is sharp, relative to the class $\mathcal{G}'(\phi, \alpha)$ on which uniformity holds. We do not know whether (34) or the rate in (b) are optimal for Theorem 5.

REMARK 6. Of course, there might exist other, entirely different, estimators behaving better than the $\widehat{\sigma}_n(k)$'s and perhaps having a better rate than in part (b) of these theorems [the rate cannot be improved in (a), of course]. We think this doubtful, however.

REMARK 7. The most interesting situation is when we have asymptotic efficiency in the sense of having the same asymptotic variance as when Y is absent (this happens when G is symmetrical), or at least "rate efficiency" (that is rate $\sqrt{n}$). We have this under (33) or (34), that is, when $\Delta_n \to 0$ sufficiently fast.

When $\Delta_n = 1/n$, rate efficiency is satisfied provided $\alpha \le \beta/2$ for Theorem 4 and provided $\alpha \le 2\beta/(4+\beta)$ for Theorem 5. If Y is a compound Poisson process with drift, then rate efficiency holds provided $n\Delta_n^2$ is bounded, regardless of the value taken by $\beta \in (0, 2]$ (take $\alpha = 0$).

REMARK 8. When we do not know that G is symmetrical, we have rate efficiency when $\Delta_n \to 0$ sufficiently fast, but cannot achieve the asymptotically efficient variance, even under (34). However, the asymptotic variances in the two theorems above are the same, up to the factor $1 - \delta$. By choosing δ small, one can approach asymptotic efficiency as closely as desired.

7. Example: the empirical characteristic function. We now turn to specific estimators. Recall that a way of estimating a parameter for i.i.d. variables X_j is to use the empirical characteristic function $\sum_{j\in J} \exp(iwX_j)$, or $\sum_{j\in J}\cos(wX_j)$ in the symmetrical case, for some given w (or several w's at once), where J is the index set; in the Lévy process setting, see, for example, [8, 10, 18, 19] and Chapter 4 in [23].

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

$$V_n(w) = \frac{1}{p_n} \sum_{i=m_n+1}^{n} \cos\biggl(\frac{w\chi_i'^n(G)}{S_n(G)}\biggr), \tag{35}$$

where $S_n(G)$ is the preliminary estimator. In other words, we take $k(x) = \cos(wx)$ [a bounded function, so $k_n = k$ in (22)] and the estimating function of (24) is

$$U_{n,G,\beta,k}(u) = V_n(w) - \Psi_{G,\Delta_n,\beta,k}\biggl(\frac{u}{S_n(G)}, \frac{1}{S_n(G)}, 0\biggr). \tag{36}$$

Furthermore, if $\rho(u)$ is the exponent in (2), we get, for $g(x) = \exp(iwx)$,

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

Taking the real part and using (13) and the fact that $G \in \mathcal{G}_\beta$, we see that for $k(x) = \cos(wx)$, we have $\Psi_{G,\Delta,\beta,k}(u, v, 0) = \exp(A_\Delta(u, v))\cos(B_\Delta(u, v))$, where

$$\begin{aligned} A_\Delta(u,v) &= -\frac{w^\beta u^\beta}{2} + \int F(dx)\bigl(\cos(wv\Delta^{1-1/\beta}x) - 1\bigr), \\ B_\Delta(u,v) &= \begin{cases} \displaystyle\int F(dx)\sin(wv\Delta^{1-1/\beta}x), & \text{if } \beta < 1, \\ \displaystyle\int F(dx)\bigl(\sin(wv\Delta^{1-1/\beta}x) - wv\Delta^{1-1/\beta}x1_{\{|x|\le 1\}}\bigr), & \text{if } \beta \ge 1. \end{cases} \end{aligned} \tag{37}$$

So we can insert these formulas directly into (36). Moreover, we have $\Psi_k(u,0) = e^{-w^\beta u^\beta/2}$. Then inserting this into (26) [recall $\Psi_k(u,0) = E(k(uW_1))$, hence $I(k) = -\Psi_k'(1,0)$ and $J(k) = (\Psi_k(2,0)+1)/2 - \Psi_k(1,0)^2$], we get

$$\Sigma^2(k) = 2\frac{1 + e^{-(2w)^\beta/2} - 2e^{-w^\beta}}{\beta^2 w^{2\beta} e^{-w^\beta}}. \tag{38}$$

When $\beta < 2$, the minimal variance is achieved for some value $w = w_\beta \in (0,\infty)$; when $\beta = 2$, the variance $\Sigma^2(k)$ goes to $1/2$ as $w \to 0$—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 (30) becomes

$$U_{n,G,\beta,k}(u) = V_n(w) - \Psi_k(u/S_n, 0), \tag{39}$$

provided that in (35), we sum over $i \in \{q_n + m_n + 1, \dots, n\}$. Moreover, $u \mapsto \Psi_k(u,0)$ is invertible, so the estimator $\widehat{\sigma}_n(k)$ takes the simple explicit form

$$\widehat{\sigma}_n(k) = S_n \frac{2^{1/\beta}}{w}\left(-\log\left(\frac{1}{p_n}\sum_{i=q_n+m_n+1}^{n} \cos\left(\frac{w\Delta_n^{-1/\beta}(\chi_i^n - B_n)}{S_n}\right)\right)\right)^{1/\beta} \tag{40}$$

if the argument of the logarithm is positive [otherwise, set, e.g., $\widehat{\sigma}_n(k) = 1$].

8. Example: power and truncated power functions. Another natural choice for the function k is a power function $k(x) = |x|^r$ 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. However, the asymptotic variance is explicit. If $m_r = E(|W_1|^r)$, we get $I(k) = -rm_r$ and $J(k) = m_{2r} - m_r^2$, hence,

$$\Sigma^2(k) = \frac{m_{2r} - m_r^2}{r^2 m_r^2}. \tag{41}$$

When $\beta = 2$, $\Sigma^2(k)$ achieves its minimum of $1/2$ (the optimal variance) at $r = 2$. When $\beta < 2$, $\Sigma^2(k)$ goes to ∞ when r increases to $\beta/2$. We conjecture that $\Sigma^2(k)$ is monotone increasing in r (this holds 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 must be bounded. So, we must "truncate" the argument, using the function

$$k_\gamma(x) = |x|^r 1_{\{|x| \le \gamma\}} \tag{42}$$

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

$$V_n(\gamma) = \frac{1}{p_n \Delta_n^{r/\beta}} \sum_{i=m_n+1}^{n} |\chi_i^n - B_n|^r 1_{\{|\chi_i^n| \le \gamma \Delta^{1/\beta}\}}, \tag{43}$$

the estimator $\widehat{\sigma}_n(k_\gamma)$ is defined by $\widehat{\sigma}_n(k_\gamma) = S_n \Psi_{k_\gamma}^{-1}(V_n(\gamma S_n)/S_n^r)$ when the argument of $\Psi_{k_\gamma}^{-1}$ above is in I' and $\widehat{\sigma}_n(k_\gamma) = 1$ (for example) otherwise. This is almost as explicit as (40) and

$$\Sigma^2(k_\gamma) = \frac{M_{\gamma,2r} - M_{\gamma,r}^2}{(r M_{\gamma,r} - 2h_\beta(\gamma)\gamma^{r+1})^2}, \qquad \text{where } M_{\gamma,s} = E(|W_1|^r 1_{\{|W_1| \le \gamma\}}). \tag{44}$$

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

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

9. 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 $\lambda > 0$. We consider a number of choices of k based on the power or truncated power variations

$$V_n(c, \kappa) = \frac{1}{p_n \Delta_n^{r/2}} \sum_{i=m_n+1}^{n} |\chi_i^n|^r 1_{\{|\chi_i^n| \le c\Delta_n^{1/2+\kappa}\}} \tag{45}$$

for $r \in (0, 2]$. The truncation rate is $c\Delta^{1/2+\kappa}$ with $c \in (0, \infty]$ and $\kappa \in (-1/2, \infty)$ (so $c = \infty$ corresponds to no truncation at all). Note that V_n above is $V_n(c\Delta_n^\kappa)$ of (43); Y is symmetrical in this model, so $B_n = 0$. The associated estimator is given by $\widehat{\sigma}_n = S_n H_{\Delta_n}^{-1}(V_n(cS_n, \kappa)/S_n^r)$, where H_Δ^{-1} is the local inverse around 1 of the function $H_\Delta(u) = E(|uW_\Delta|^r 1_{\{|uW_\Delta| \le c\Delta^{1/2+\kappa}\}})$; the preliminary estimator S_n is not needed in this case and one could set $S_n = 1$. When $c = \infty$, we get the (nontruncated) rth power variation. If $c < \infty$ and $\kappa = 0$, this corresponds to taking $k = k_c$, as given by (42).

The expected values of the truncated power variations are available in closed form in this model, using the incomplete Gamma function of order a, which we denote by $\Gamma(a, \cdot)$. As described in the general theory, in the semiparametric case, we use an approximate centering based on expectations of, in this case, truncated moments, computed for the model $X = \sigma W$ (without Y). The effect of the misspecification is to bias the resulting estimator. 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

$$(46)\quad \sqrt{n\Delta_n^{v_1}}(\widehat{\sigma}_n - \overline{\sigma}_n) \to N(0, v_0), \qquad \text{where } \overline{\sigma}_n = \sigma + b_0\Delta_n^{b_1} + o(\Delta_n^{b_1})$$

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 rate of convergence slower than $n^{1/2}$. 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 carried out under the assumption that Y is present, instead of the approximate centering done without Y. 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 fully characterize the asymptotic distribution of the semiparametric estimator of σ; that is, we characterize the values (b_0, b_1, v_0, v_1) in (46) as functions of (r, κ, c) and (σ, η, λ).

9.1. *Power variations without truncation.* If $c = \infty$, we have the following for the asymptotic variance:

- if $0 < r < 1$, then $v_1 = 0$ and $v_0 = \frac{1}{r^2}(\sqrt{\pi}\frac{\Gamma(\frac{1}{2}+r)}{\Gamma(\frac{1+r}{2})^2} - 1)$;
- if $r = 1$, then $v_1 = 0$ and $v_0 = \frac{1}{2}((\pi - 2)\sigma^2 + \pi\lambda\eta)$;
- if $1 < r < 2$, then $v_1 = r - 1$ and $v_0 = \frac{\sqrt{\pi}\sigma^{2-2r}\lambda\eta^r}{r^2}\frac{\Gamma(\frac{1}{2}+r)}{\Gamma(\frac{1+r}{2})^2}$.

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

REMARK 9. The estimators based on power variations converge (not taking the bias into consideration) at rate $n^{1/2}$ only when $r \le 1$. 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 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.

REMARK 10. When $r < 1$, the asymptotic variance v_0 is identical to the expression obtained without jumps, as is the case when the log-likelihood score is 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.

9.2. *Power variations with* $\Delta^{1/2}$ *truncation.* If $c < \infty$ and $\kappa = 0$, then $v_1 = 0$ for all $r \in (0, 2]$ and

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

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

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

REMARK 11. 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 that which results from truncating at rate $\Delta_n^{1/2}$, depending on the values of (σ^2, c).

REMARK 12. 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 (this type of truncation leads to the estimators studied in Section 8). In all cases, the bias is smaller than that which we would have without truncation.

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

- if $-3/(2+4r) < \kappa < 0$, then $v_1 = 0$ and $v_0 = \frac{\sigma^2}{r^2}(\sqrt{\pi}\frac{\Gamma(\frac{1}{2}+r)}{\Gamma(\frac{1+r}{2})^2} - 1)$;
- if $\kappa = -3/(2+4r)$, then $v_1 = 0$ and

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

- if $-1/2 < \kappa < -3/(2+4r)$, then $v_1 = -\kappa - 2r\kappa - 3/2 > 0$ 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 the following:

- if $-1/(2+2r) < \kappa < 0$, then $b_1 = 1$ and $b_0 = -\frac{\lambda\sigma}{r}$;
- if $\kappa = -1/(2+2r)$, then $b_1 = 1$ and $b_0 = \frac{\lambda\sigma}{(1+r)}(\frac{2^{1/2-r/2}c^{1+r}}{r\sqrt{\eta}\sigma^r\Gamma(\frac{1+r}{2})} - 1 - \frac{1}{r})$;
- if $-1/2 < \kappa < -1/(2+2r)$, then $b_1 = 3/2 + \kappa + r\kappa > 0$ 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 13. When $0 < r < 1$, we are automatically in the situation where $\kappa > -3/(2+4r)$, 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 that which results from keeping all increments (i.e., not truncating at all). When $1 < r < 2$, it is possible to obtain the convergence rate $n^{1/2}$ by keeping more than $O(\Delta_n^{1/2})$ increments, but still "not too many" of them ($-3/(2+4r) \le \kappa < 0$); 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 $\widehat{\sigma}_n$, although slower than $n^{1/2}$, is nonetheless faster than $n^{1/2}\Delta_n^{(r-1)/2}$.

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

$$v_0 = \frac{\sqrt{2\pi}(1+r)^2\sigma^3}{2c(1+2r)}, \qquad b_1 = 1, \qquad b_0 = \sigma\lambda.$$

REMARK 14. Truncating at a rate faster than $\Delta^{1/2}$ causes the convergence rate of the estimator to deteriorate 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 (by truncating "too much"), which increases its asymptotic variance.

9.5. *Comparison with the general case.* Let us compare the specific results just obtained for this particular model with the general results obtained in Theorems 4 and 5. Here, we have $G \in \overline{\mathcal{G}}_0'$, so the general results assert that if

$$n\Delta_n^2 \to 0, \tag{47}$$

then the estimators $\widehat{\sigma}_n$ converge at a rate $\sqrt{n}$ and the limit of the normalized error is Gaussian without bias; when (47) fails, but $\Delta_n \to 0$ still holds, the sequence $(\sqrt{n} \wedge \Delta_n^{-1})(\widehat{\sigma}_n - \sigma)$ is tight. The estimators converge at rate $\sqrt{n}$ when $v_1 = 0$ and $n\Delta_n^{2b_1}$ is bounded (then there is a bias) or when $n\Delta_n^{2b_1} \to 0$ (there is no bias). Otherwise, the sequence $(\sqrt{n\Delta_n^{v_1}} \wedge \Delta_n^{-b_1})(\widehat{\sigma}_n - \sigma)$ is tight. Then we have the following:

- 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 worse than in our general results (this was expected, of course).
- Power variation with $\Delta^{1/2}$ truncation: If $n\Delta_n^2 \to 0$, then we have rate $\sqrt{n}$ with asymptotically unbiased error. If $n\Delta_n^2 \to a \in (0, \infty)$, we have rate $\sqrt{n}$ with asymptotically biased error. If $n\Delta_n^2 \to \infty$, then $\Delta_n^{-1}(\widehat{\sigma}_n - \sigma)$ converges in probability to the constant b_0, this being slightly better than what we get by applying the general results. This holds irrespective of r, but, of course, the asymptotic variance depends on r and also on c.

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

10. Example: sum of two stable processes. We now suppose that Y is also a symmetric stable process with index $\alpha \in (0, \beta)$, so that $G \in \overline{\mathcal{G}}'_{\alpha}$. Related results for this model can be found in [13]. First, we can consider estimators based on the empirical characteristic function, that is, $k(x) = \cos(wx)$ for some $w > 0$. We have the parametric estimator $\widehat{\sigma}_n = \widehat{\sigma}_n(G, \phi, k)$ of (25) and the sequence $\sqrt{n}(\widehat{\sigma}_n - \sigma)$ converges in law to $N(0, \sigma^2\Sigma^2(k))$, where $\Sigma^2(k)$ is given by (38). We also have the semiparametric estimators $\widehat{\sigma}_n(k)$ of (31), which behave as follows. Under

$$n\Delta_n^{2(\beta-\alpha)/\beta} \to 0, \tag{48}$$

$\sqrt{n}(\widehat{\sigma}_n(k) - \sigma)$ converges in law to $N(0, \sigma^2\Sigma^2(k))$ and the sequence $(\sqrt{n} \wedge \Delta_n^{-(\beta-\alpha)/\beta})\,(\widehat{\sigma}_n - \sigma)$ is always tight. In fact, since we are in the symmetrical case, the preliminary estimator $S_n = S_n(G)$ is the same for both $\widehat{\sigma}_n$ and $\widehat{\sigma}_n(k)$, which are the solutions of $U_n(u) = 0$ and $U'_n(u) = 0$, respectively, closest to S_n and

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

[recall (36) and (39)]. If we use the explicit form (37), we obtain

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

which is equivalent to $(w^\alpha/2\sigma^\alpha)\Delta_n^{(\beta-\alpha)/\beta}\exp(-w^\beta/2)$ as $n \to \infty$ and $u \to \sigma$ (recall that $S_n \to \sigma$ in probability). Since $\Psi_k(u, 0) = \exp(-u^\beta w^\beta/2)$, we have $\partial\Psi_k(1,0)/\partial u = -\beta w^\beta \exp(-w^\beta/2)/2 \neq 0$ and deduce that the difference $\widehat{\sigma}_n(k) - \widehat{\sigma}_n$ is equivalent (in probability) to $-\Delta_n^{(\beta-\alpha)/\beta}/(w^{\beta-\alpha}\beta\sigma^\alpha)$.

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 (48), we have the following:

- if $n\Delta_n^{2(\beta-\alpha)/\beta} \to a^2 \in (0, \infty)$, then $\sqrt{n}(\widehat{\sigma}_n(k) - \sigma)$ converges in law to $N(-a/(w^{\beta-\alpha}\beta\sigma^\alpha), \sigma^2\Sigma^2(k))$;
- if $n\Delta_n^{2(\beta-\alpha)/\beta} \to \infty$, then $\Delta_n^{-(\beta-\alpha)/\beta}(\widehat{\sigma}_n(k) - \sigma)$ converges in probability to the constant $-1/(w^{\beta-\alpha}\beta\sigma^\alpha)$.

We conclude that the results of Theorem 4 are sharp, at least for $k(x) = \cos(wx)$.

We can undertake a similar analysis for the estimators based on the truncated power variation $V_n(\gamma)$ of (43) with $B_n = 0$ (because Y is symmetrical). That is, we consider the truncated power variations at the level $\Delta_n^{1/\beta}$. When $n\Delta_n^{2(\beta-\alpha)/\beta} \to \infty$, one can show that for sufficiently small γ (but it is probably true for all $\gamma > 0$), the sequence $\Delta_n^{-(\beta-\alpha)/\beta}\,(\widehat{\sigma}_n - \sigma)$ is tight and its limiting distributions include some Dirac masses at nonvanishing constants. So, here, again, the results of Theorem 4 are sharp. But, of course, as previously stated, this does not completely rule out the existence of estimators constructed in a different way and behaving better.

11. Technical preliminaries. In the sequel, 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.

LEMMA 1. *Let $\phi \in \overline{\Phi}$, $\alpha \in (0, 2]$ and ϕ_α be the associated function defined as in* (17). *For all $G \in \mathcal{G}(\phi, \alpha)$ and $\varepsilon \in (0, 1]$, we have*

$$\int_{\{|x|\le\varepsilon\}} |x|^q F(dx) \le \begin{cases} \dfrac{q}{q-\alpha}\varepsilon^{q-\alpha}\phi_\alpha(\varepsilon), & \text{if } q > \alpha, \\ \phi_\alpha(\varepsilon), & \text{if } q = \alpha = 2, \end{cases} \tag{49}$$

$$\int_{\{\varepsilon<|x|\le 1\}} |x| F(dx) \le \begin{cases} \phi_\alpha(1), & \text{if } \alpha < 1, \\ \phi_\alpha(\varepsilon)\log(1/\varepsilon), & \text{if } \alpha = 1, \\ \phi_\alpha(\varepsilon)\varepsilon^{1-\alpha}, & \text{if } \alpha > 1. \end{cases} \tag{50}$$

PROOF. When $q = \alpha = 2$, (49) is trivial because $\phi \le \phi_\alpha$. When $q > \alpha$, Fubini's theorem and (3) together yield

$$\begin{aligned} \int_{\{|x|\le\varepsilon\}} |x|^q F(dx) &= \int_{\{|x|\le\varepsilon\}} F(dx) q \int_0^{|x|} y^{q-1}\,dy = q\int_0^\varepsilon y^{q-1} F(|x|>y)\,dy \\ &\le q\int_0^\varepsilon \phi(y) y^{q-1-\alpha}\,dy \le \frac{q}{q-\alpha}\phi'(\varepsilon)\varepsilon^{q-\alpha} \end{aligned}$$

because ϕ' is increasing [recall the notation in (17)]. So, we again get (49). In a similar way, for every $z \in [\varepsilon, 1]$, we get

$$\begin{aligned} \int_{\{\varepsilon<|x|\le 1\}} |x| F(dx) &= \int_{\{\varepsilon<|x|\le 1\}} F(dx) \int_0^{|x|} dy \\ &= \int_0^\varepsilon F(\varepsilon < |x| \le 1)\,dy + \int_\varepsilon^z F(y < |x| \le 1)\,dy \\ &\quad + \int_z^1 F(y < |x| \le 1)\,dy \\ &\le \phi'(\varepsilon)\varepsilon^{1-\alpha} + \phi'(z)\int_\varepsilon^z y^{-\alpha}\,dy + \phi'(1)\int_z^1 y^\alpha\,dy. \end{aligned}$$

A simple calculation, using (17), allows us to deduce (50); we take $z = 1$ when $\alpha < 1$, $z = 1$ when $\alpha = 1$ and $\varepsilon \ge 1/e$, $z = \exp -\sqrt{\log(1/\varepsilon)}$ when $\alpha = 1$ and $\varepsilon < 1/e$ and $z = \sqrt{\varepsilon}$ when $\alpha > 1$. $\square$

In the next lemma, we use the notation $b'(G,\alpha)$, $Z_\Delta(\alpha)$ and $G'_{\Delta,\alpha}$, introduced in (13) and thereafter, and also $\rho(\alpha,\beta)$ and $\rho'(\alpha,\beta)$, as defined by (32).

LEMMA 2. (a) *If $G \in \mathcal{G}_\beta$, then $G'_{\Delta,\beta}$ converges to the Dirac mass δ_0 as $\Delta \to 0$.*

(b) *If $\alpha \le \beta$, $\phi \in \Phi$, G^n is a sequence of measures in $\mathcal{G}(\phi,\alpha)$ and $\Delta_n \to 0$, then the associated sequence $G'^n_{\Delta_n,\alpha}$ converges to the Dirac mass δ_0 as $n \to \infty$.*

(c) *If $\alpha \le \beta$ and $\phi \in \overline{\Phi}$, then there exists a constant $C = C_{\phi,\alpha}$ such that for all functions g with $|g(x)| \le K(1 \wedge |x|)$ and all $\Delta \in (0,1]$, we have* (*with ϕ_α defined as in the previous lemma*)

$$
\begin{aligned}
&G \in \mathcal{G}(\phi,\alpha), |g(x)| \le K(1 \wedge |x|) \\
&\qquad \Longrightarrow \quad E(|g(Z_\Delta(\alpha))|) \le CK\Delta^{\rho(\alpha,\beta)}\phi_\alpha\big(\Delta^{\frac{2+\beta}{\beta(2+\alpha)}}\big), \\
&G \in \mathcal{G}'(\phi,\alpha), |g(x)| \le K(1 \wedge x^2) \\
&\qquad \Longrightarrow \quad E(|g(Z_\Delta(\alpha))|) \le CK\Delta^{\rho'(\alpha,\beta)}\phi_\alpha\big(\Delta^{\frac{1}{\beta}}\big).
\end{aligned}
\tag{51}
$$

PROOF. If $\phi \in \Phi$, then $\lim_{x\to 0}\phi_\alpha(x) = 0$, so (c)$\Rightarrow$(b)$\Rightarrow$(a). For proving (c), we say that we are in the asymmetrical (resp. symmetrical) case if G and g are given as in the first (resp. the second) statement in (51).

Let $\eta \in (0,1/2]$, to be chosen later. With any given $G \in \mathcal{G}(\phi,\alpha)$, we associate the Lévy process Y and the characteristics $(b,0,F)$. Let F' and F'' be the restrictions of F to the sets $[-\eta,\eta]$ and $[-\eta,\eta]^c$, respectively. We can decompose Y into the sum $Y_t = at + Y'_t + Y''_t$, where Y' is a Lévy process with characteristics $(0,0,F')$, Y'' is a compound Poisson process with Lévy measure F'' and $a = b - \int_{\{\eta<|x|\le 1\}} xF(dx)$. Then $a' = a - b'(G,\alpha)$ is given by

$$
a' = \begin{cases} \displaystyle\int_{\{|x|\le\eta\}} xF(dx), & \text{if } \alpha < 1, \\ \displaystyle -\int_{\{\eta<|x|\le 1\}} xF(dx), & \text{if } \alpha \ge 1. \end{cases}
$$

Therefore, $a' = 0$ in the symmetrical case, whereas in the asymmetrical case, (49) and (50) yield for a constant $C = C_\alpha$ not depending on $G \in \mathcal{G}(\phi,\alpha)$,

$$
|a'| \le \begin{cases} C\eta^{1-\alpha}\phi_\alpha(\eta), & \text{if } \alpha \ne 1, \\ C\log(1/\eta)\phi_\alpha(\eta), & \text{if } \alpha = 1. \end{cases}
\tag{52}
$$

Also, since Y' has no drift, no Wiener part and no jump bigger than 1, one knows [by differentiating (2), for example] that $E((Y'_t)^2) = t\int x^2F'(dx)$. Then (49) again yields, for some $C = C_\alpha$,

$$
E(|Y'_\Delta|^2) \le C\Delta\eta^{2-\alpha}\phi_\alpha(\eta).
\tag{53}
$$

We have $|g(Z_\Delta(\alpha))| \le K$. If, further, $Y''_\Delta = 0$, then $Y_\Delta = a\Delta + Y'_\Delta$, therefore $Z_\Delta(\alpha) = \Delta^{-1/\beta}(Y'_\Delta + a'\Delta)$, therefore $|g(Z_\Delta(\alpha))| \le K\Delta^{-1/\beta}(|Y'_\Delta| + \Delta|a'|)$ in the asymmetrical case and $|g(Z_\Delta(\alpha))| \le K\Delta^{-2/\beta}Y'^2_\Delta$ in the symmetrical case. Next, $P(Y''_\Delta \ne 0) \le \Delta F''(R) \le \Delta\phi_\alpha(\eta)/\eta^\alpha$ because $G \in \mathcal{G}(\phi,\alpha)$. Therefore, we deduce from (52) and (53) that for some constant $C = C_{\phi,\alpha}$,

$$E(|g(Z_\Delta(\alpha))|) \le \begin{cases} CK\bigl(\Delta\eta^{-1} + \Delta^{1/2-1/\beta}\eta^{1/2} + \Delta^{1-1/\beta}\log(1/\eta)\bigr)\phi_1(\eta), & \text{if } \alpha = 1,\\ CK\bigl(\Delta\eta^{-\alpha} + \Delta^{1/2-1/\beta}\eta^{1-\alpha/2} + \Delta^{1-1/\beta}\eta^{1-\alpha}\bigr)\phi_\alpha(\eta), & \text{otherwise}\end{cases}$$

in the asymmetrical case and $E(|g(Z_\Delta(\alpha))|) \le CK\bigl(\Delta\eta^{-\alpha} + \Delta^{1-2/\beta}\eta^{2-\alpha}\bigr)\phi_\alpha(\eta)$ otherwise. Then take $\eta = \Delta^{(2+\beta)/\beta(2+\alpha)}$ (resp. $\eta = \Delta^{1/\beta}$) to obtain (51). □

Next, we study the functions defined in (15).

LEMMA 3. (a) *Let k satisfy the first half of (20) with some $\gamma \ge 0$ such that $\gamma < \beta$ whenever $\beta < 2$. Then Ψ_k is C^∞ on $(0,\infty)\times\mathbb{R}$. If, in addition, $\gamma > 0$, $\nu \in (0,\infty)$ and $k_\nu(x) = k(x)1_{\{|k(x)|\le\nu\}}$, then for all $K > 0$, there exists a constant $M_{K,k}$ such that $|z| \le K$ and $\nu \ge M_{K,k}$ imply*

$$\left|\frac{\partial^{j+l}}{\partial u^j\partial z^l}\Psi_k(u,z) - \frac{\partial^{j+l}}{\partial u^j\partial z^l}\Psi_{k_\nu}(u,z)\right| \le \begin{cases} C_{j,l,k,K}u^{\beta-j}\nu^{1-(l+\beta)/\gamma}, & \text{if } \beta < 2,\\ C_{j,l,k,K}u^\gamma e^{-\nu^{1/\gamma}/u}, & \text{if } \beta = 2.\end{cases} \tag{54}$$

(b) *If k is bounded, then for all $\eta \in (0,1)$, we have, with $\|k\|$ being the sup norm,*

$$\eta \le u \le 1/\eta \quad\Longrightarrow\quad \left|\frac{\partial^{j+l}}{\partial u^j\partial z^l}\Psi_k(u,z)\right| \le C_{l,j,\eta}\|k\|. \tag{55}$$

PROOF. We can rewrite the second display of (15) as

$$\Psi_k(u,z) = \frac{1}{u}\int h_\beta\left(\frac{x}{u}\right)k(x+z)\,dx = \frac{1}{u}\int h_\beta\left(\frac{x-z}{u}\right)k(x)\,dx. \tag{56}$$

(a) For an integer l, the jth 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 $h_{l,j}(x)$ which is a linear combination of products of $h_\beta^{(l+i)}(x)x^i$ for $0 \le i \le j$. By (7), $h_{l,j}$ satisfies, for all x,

$$|h_{l,j}(x)| \le \begin{cases} C_{j,l}/(1+|x|^{1+l+\beta}), & \text{if } \beta < 2,\\ C_{j,l}(1+|x|^{2j+2l})e^{-x^2/2}, & \text{if } \beta = 2.\end{cases} \tag{57}$$

The above estimate for $\beta < 2$ also holds for $\beta = 2$. Moreover, for all $\beta \in (0, 2]$, the derivative satisfies

$$|h_{l,j}^{(1)}(x)| \le \frac{C_{j,l}}{1 + |x|^{2+l+\beta}}. \tag{58}$$

Therefore, we easily deduce from (56) that Ψ_k is C^∞, with [by differentiating l times the last term in (56), then differentiating j times the analogue of the last term, with $h_\beta^{(l)}$ instead of h_β]

$$\begin{aligned} \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_k(u,z) &= \frac{1}{u^{j+l+1}} \int h_{l,j}(x/u) k(x+z)\, dx \\ &= \frac{1}{u^{j+l}} \int h_{l,j}(x) k(ux+z)\, dx. \end{aligned} \tag{59}$$

In particular, since $|k(x) - k_\nu(x)| = |k(x) 1_{\{|k(x)| > \nu\}}| \le C_k (1 + |x|^\gamma) 1_{\{|k(x)|>\nu\}}$, we have

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

since $\{(x,z) : |k(x+z)| > \nu\} \subset \{(x,z) : 1 + |x+z|^\gamma > \nu\varepsilon_k\}$ for some $\varepsilon_k > 0$. Then a simple computation using (57) yields (54).

(b) When k is bounded, (57) implies that $h_{l,j}$ is integrable and so (59) yields (55). □

LEMMA 4. *If k is bounded, then $\Psi_{G,\Delta,\alpha,k}(u,v,z)$ is C^∞ in (u,z) and for any $\eta \in (0,1)$, we have*

$$\eta \le u \le 1/\eta \quad \Longrightarrow \quad \left| \frac{\partial^{j+l}}{\partial u^j \partial z^l} \Psi_{G,\Delta,\alpha,k}(u,v,z) \right| \le C_{l,j,\eta} \|k\| \tag{60}$$

and also, for $\Delta \le 1$, and $z \in \mathbb{R}$, $u \in [\eta, 1/\eta]$ and $v \in (0, 1/\eta]$,

$$\begin{aligned} G \in \mathcal{G}(\phi,\alpha) \Longrightarrow &\left| \frac{\partial^j}{\partial u^j} \Psi_{G,\Delta,\alpha,k}(u,v,z) - \frac{\partial^j}{\partial u^j} \Psi_k(u,0) \right| \\ &\qquad \le C_{j,\eta} \|k\| \big(|z| + \Delta^{\rho(\alpha,\beta)} \phi_\alpha\big(\Delta^{\frac{2+\beta}{\beta(2+\alpha)}}\big)\big), \\ G \in \mathcal{G}'(\phi,\alpha) \Longrightarrow &\left| \frac{\partial^j}{\partial u^j} \Psi_{G,\Delta,\alpha,k}(u,v,z) - \frac{\partial^j}{\partial u^j} \Psi_k(u,0) \right| \\ &\qquad \le C_{j,\eta} \|k\| \big(|z| + \Delta^{\rho'(\alpha,\beta)} \phi_\alpha\big(\Delta^{\frac{1}{\beta}}\big)\big). \end{aligned} \tag{61}$$

PROOF. Observe that by (15), $\Psi_{G,\Delta,\alpha,k}(u,v,z)=\int G'_{\Delta,\alpha}(dw)\Psi_k(u,vw+z)$. Then by (55), $\Psi_{G,\Delta,\alpha,k}$ is C^∞ in (u,z), with

$$\frac{\partial^{j+l}}{\partial u^j\partial z^l}\Psi_{G,\Delta,\alpha,k}(u,v,z)=\int G'_{\Delta,\alpha}(dw)\frac{\partial^{j+l}}{\partial u^j\partial z^l}\Psi_k(u,vw+z) \tag{62}$$

and for any $\eta\in(0,1)$, we have (60). Next, we prove the first part of (61). From (58), we have that

$$|y|\le 1 \quad\Longrightarrow\quad |h_{0,j}(x+y)-h_{0,j}(x)|\le C_{j,m}\frac{|y|}{1+|x|^{2+\beta}}. \tag{63}$$

Recalling (59) and (62), we have

$$\frac{\partial^j}{\partial u^j}\Psi_{G,\Delta,\alpha,k}(u,v,z)-\frac{\partial^j}{\partial u^j}\Psi_k(u,z)=\int G'_{\Delta,\alpha}(dw)g(w), \tag{64}$$

where

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

for u,v,z,j fixed. Let $\eta\in(0,1)$ and suppose that $\eta\le u\le 1/\eta$ and $v\le 1/\eta$. If $|w|\le 1$, then (63) yields $|g(w)|\le C_{j,\eta}\|k\||w|$, whereas (55) always yields $|g(w)|\le C_{j,\eta}\|k\|$; so, $|g(w)|\le C_{j,\eta}\|k\|(|w|\wedge 1)$ and in view of (64), we readily deduce from (51) that

$$\left|\frac{\partial^j}{\partial u^j}\Psi_{G,\Delta,\alpha,k}(u,v,z)-\frac{\partial^j}{\partial u^j}\Psi_k(u,z)\right|\le C_{j,\eta}\|k\|\Delta^{\rho(\alpha,\beta)}\phi_\alpha\big(\Delta^{\frac{2+\beta}{\beta(2+\alpha)}}\big). \tag{65}$$

Since, further, $|\frac{\partial^j}{\partial u^j}\Psi_k(u,z)-\frac{\partial^j}{\partial u^j}\Psi_k(u,0)|\le C_{j,\eta}\|k\||z|$ by (55), we obtain the result.

Finally, the function $h_{0,j}$ is C^∞ and all of its derivatives satisfy the estimates (57). Hence, the functions $h^{(1)}_{0,j}$ and $H(x)=\sup_{y\in[x-1/\eta^2,x+1/\eta^2]}|h^{(2)}_{0,j}(y)|$ are integrable. We have

$$|w|\le 1\quad\Rightarrow\quad\left|h_{0,j}\left(x-\frac{vw}{u}\right)-h_{0,j}(x)-h^{(1)}_{0,j}(x)\frac{vw}{u}\right|\le C_{j,\eta}w^2H(x), \tag{66}$$

provided that $v<1/\eta$ and $\eta\le u\le 1/\eta$. Therefore, we can write $g=g_1+g_2$, where

$$\begin{aligned} g_1(w)&=\frac{vw}{u^{j+1}}1_{\{|w|\le 1\}}\int h^{(1)}_{0,j}(x)k(ux+z)\,dx,\\ g_2(w)&=g(w)1_{\{|w|>1\}}\\ &\quad+1_{\{|w|\le 1\}}\int\left(h_{0,j}\left(x-\frac{vw}{u}\right)-h_{0,j}(x)-h^{(1)}_{0,j}(x)\frac{vw}{u}\right)k(ux+z)\,dx. \end{aligned}$$

On the one hand, if $G \in \mathcal{G}'(\phi, \alpha)$, then $G'_{\Delta,\alpha}$ is symmetrical about 0 and $\int g_1(w) G'_{\Delta,\alpha}(dw) = 0$ because g_2 is bounded and odd. On the other hand, (66), the integrability of H and $|g(w)| \le C_{j,\eta}\|k\|$ collectively yield $|g_2(w)| \le C_{j,\eta}\|k\|(w^2 \wedge 1)$. Hence, using (51), we get [instead of (65)]

$$\left|\frac{\partial^j}{\partial u^j}\Psi_{G,\Delta,\alpha,k}(u,v,z) - \frac{\partial^j}{\partial u^j}\Psi_k(u,z)\right| \le C_{j,\eta}\|k\|\Delta^{\rho'(\alpha,\beta)}\phi_\alpha(\Delta^{1/\beta})$$

and the second part of (61) follows analogously. □

12. Estimating equations. We prove here a general result concerning estimating equations, which will be used several times below. Its content is basically known, but we adapt it to our setting, at the desired level of generality.

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

(A1) if $\sigma_n \to \sigma > 0$, then $S_n \to \sigma$ in P_{n,σ_n}-probability;
(A2) if $\sigma_n \to \sigma > 0$, then the sequence $(Q_n \mid P_{n,\sigma_n})$ is tight.

Next, we consider two families of functions $(f_{n,s,q})_{s>0}$ and $(H_{n,s})_{s>0,q\in\mathbb{R}}$, on $\mathbb{R}$ and $(0,\infty)$, respectively, and we associate with them the following estimating functions and estimators:

$$U_{n,s,q}(u) = \frac{1}{p_n}\sum_{i=1}^{p_n}\big(f_{n,s,q}(\chi_i^n) - H_{n,s}(u)\big), \tag{67}$$

$$\widehat{\sigma}_n(s,q) = \begin{cases} \text{the } u > 0 \text{ with } U_{n,s,q}(u) = 0 \text{ which is closest to } s, \\ \qquad\qquad \text{if it exists,} \\ 1, \qquad\quad \text{otherwise} \end{cases} \tag{68}$$

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

$$F_{n,s,q}(\sigma) = E_{n,\sigma}(f_{n,s,q}(\chi_i^n)), \qquad F'_{n,s,q}(\sigma) = E_{n,\sigma}(f_{n,s,q}(\chi_i^n)^2). \tag{69}$$

Let us now list a series of assumptions on the previously introduced functions.

(B1) We have $\sup_{s>0,\ q\in\mathbb{R}} \|f_{n,s,q}\|^4/p_n \to 0$ when $n \to \infty$ (here, $\|f_{n,s,q}\|$ is the sup norm).
(B2) $H_{n,s}$ is continuously differentiable.
(B3) For all $s > 0$, there exists a differentiable function $\overline{F}_s$ on $(0,\infty)$ such that whenever $s_n \to s$, H_{n,s_n} and $H^{(1)}_{n,s_n}$ converge locally uniformly to $\overline{F}_s$ and $\overline{F}^{(1)}_s$, respectively.

(B4) $\overline{F}_s^{(1)}(s) \neq 0$ for all $s > 0$.
(B5) $F'_{n,s_n,q_n}(u_n)$ converges to a limit $F'(u)$ for any two sequences u_n and s_n converging to the same limit $u > 0$ and any bounded sequence q_n.
(B6) There exists a sequence $w_n \to +\infty$ 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.

THEOREM 6. *Assume* (A1), (A2) *and* (B1)–(B6).

(a) *The sequence* $((w_n \wedge \sqrt{p_n})(\widehat{\sigma}_n(S_n, Q_n) - \sigma))$ *is tight under* $P_{n,\sigma}$, *uniformly in n and in* σ *in any compact subset of* $(0, \infty)$.
(b) *If* $w_n/\sqrt{p_n} \to \infty$, *then the sequence* $(\sqrt{p_n}(\widehat{\sigma}_n(S_n, Q_n) - \sigma))$ *converges in law under* $P_{n,\sigma}$, *uniformly in* σ *in any compact subset of* $(0, \infty)$, *toward the centered normal distribution with variance* $\Xi^2(\sigma) := (F'(\sigma) - \overline{F}_\sigma(\sigma)^2)/\overline{F}_\sigma^{(1)}(\sigma)^2$.

We devote the remainder of this section to proving this theorem, assuming for the rest of the section (A1), (A2) and (B1)–(B6). First, we state a lemma which gathers some classical limit theorems. For each n, let $(\zeta_i^n : i = 1, \dots, \kappa_n)$ be real-valued and i.i.d. random variables, possibly defined on different probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ when n varies.

LEMMA 5. *Assume that* ζ_i^n *is square integrable and set* $\gamma_n = E_n(\zeta_i^n)$ *and* $\Gamma_n = E_n((\zeta_i^n)^2) - \gamma_n^2$. *If* $\kappa_n \to \infty$ *and* $\Gamma_n/\kappa_n \to 0$, *then we have*

$$\frac{1}{\kappa_n}\sum_{i=1}^{\kappa_n} \zeta_i^n - \gamma_n \xrightarrow{L^2(P_n)} 0. \tag{70}$$

Furthermore, if $\Gamma_n \to \Gamma$ *for some limit* $\Gamma \geq 0$ *and if* $E(|\zeta_i^n|^4)/\kappa_n \to 0$, *then we have*

$$\sqrt{\kappa_n}\left(\frac{1}{\kappa_n}\sum_{i=1}^{\kappa_n} \zeta_i^n - \gamma_n\right) \xrightarrow{L(P_n)} N(0, \Gamma). \tag{71}$$

In the next three lemmas, we suppose that $\sigma_n \to \sigma > 0$ and write $P_n = P_{n,\sigma_n}$.

LEMMA 6. *Let* $s_n \to \sigma$ *and let* q_n *be a bounded sequence.*

(a) *The sequence* $((w_n \wedge \sqrt{p_n})U_{n,s_n,q_n}(\sigma_n) \mid P_n)$ *is tight.*
(b) *If* $w_n/\sqrt{p_n} \to \infty$, *then the sequence* $(\sqrt{p_n}U_{n,s_n,q_n}(\sigma_n) \mid P_n)$ *converges in law to* $N(0, F'(\sigma) - \overline{F}_\sigma(\sigma)^2)$.

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

$$\gamma_n = F_{n,s_n,q_n}(\sigma_n) - H_{n,s_n}(\sigma_n), \qquad \Gamma_n = F'_{n,s_n,q_n}(\sigma_n) - F_{n,s_n,q_n}(\sigma_n)^2,$$

respectively, and, further, $|\zeta_i^n| \le \alpha_n$ for α_n satisfying $\alpha_n^4/p_n \to 0$, by (B1). Now, (B6) yields that $\gamma_n \to 0$, which, together with (B3), yields $F_{n,s_n,q_n}(\sigma_n) \to \overline{F}_\sigma(\sigma)$. On the other hand, (B5) implies $F'_{n,s_n,q_n}(\sigma_n) \to F'(\sigma)$. Therefore, $(\sqrt{p_n}(U_{n,s_n,q_n}(\sigma_n) - \gamma_n) \mid P_n)$ converges in law to $N(0, F'(\sigma) - \overline{F}_\sigma(\sigma)^2)$ by (71) and since $\sup_n w_n|\gamma_n| < \infty$ by (B6), we readily obtain the two results. □

LEMMA 7. (a) *The sequence* $((w_n \wedge \sqrt{p_n})U_{n,S_n,Q_n}(\sigma_n) \mid P_n)$ *is tight.*

(b) *If* $w_n/\sqrt{p_n} \to \infty$, *then the sequence* $(\sqrt{p_n}U_{n,S_n,Q_n}(\sigma_n) \mid P_n)$ *converges in law to* $N(0, F'(\sigma) - \overline{F}_\sigma(\sigma))$.

PROOF. For both results, it is enough to prove that from any subsequence one can extract a sub-subsequence for which the results hold and thus, using (A1) and (A2), we can indeed assume that the pair (S_n, Q_n) converges in law. Due to the independence of (S_n, Q_n) and $(U_{n,s,q} : s > 0, q \in \mathbb{R})$, we can replace the pair (S_n, Q_n) in U_{n,S_n,Q_n} by any other pair (S'_n, Q'_n) having the same law under P_n as (S_n, Q_n) and still independent of $(U_{n,s,q} : s > 0, q \in \mathbb{R})$. In particular, by the Skorokhod representation theorem, we can indeed assume that S_n converges pointwise to σ and also that Q_n converges pointwise to a limit Q.

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

$$\lim_{k\to\infty} \sup_{n\ge1} u_{k,B}(n, s_n) = 0, \qquad \text{where } u_{k,B}(n, s) = \sup_{|q|\le B} P_n(|V_{n,s,q}| > k). \tag{72}$$

If the sequence $(V_{n,S_n,Q_n} \mid P_n)$ is not tight, then there exist some $\varepsilon > 0$ and an infinite sequence n_k such that $P_{n_k}(|V_{n_k,S_{n_k},Q_{n_k}}| > k) \ge \varepsilon$. Since (S_n, Q_n) is independent of the family $(V_{n,s,q} : s > 0, q \in \mathbb{R})$, we have

$$\begin{aligned}
&P_{n_k}(|V_{n_k,S_{n_k},Q_{n_k}}| > k)\\
&\quad\le E_{n_k}[P_{n_k}(|V_{n_k,S_{n_k},Q_{n_k}}| > k|S_{n_k}, Q_{n_k})]\\
&\quad\le E_{n_k}[1_{\{|Q_{n_k}|>B\}} + 1_{\{|Q_{n_k}|\le B\}}P_{n_k}(|V_{n_k,S_{n_k},Q_{n_k}}| > k|S_{n_k}, Q_{n_k})]\\
&\quad\le P_{n_k}(|Q_{n_k}| > B) + E_{n_k}(u_{k,B}(n_k, S_{n_k}))).
\end{aligned}$$

Then by (72), $S_{n_k} \to \sigma$ and Lebesgue's theorem, we have

$$\limsup_k P_{n_k}(|V_{n_k,S_{n_k},Q_{n_k}}| > k) \le \sup_k P_{n_k}(|Q_{n_k}| > B)$$

for all $B > 0$ and, in view of (A2), we deduce that $\limsup_k P_{n_k}(|V_{n_k,S_{n_k},Q_{n_k}}| > k) = 0$. This contradicts the definition of the sequence n_k and we thus have the result.

(b) Let us denote by V a variable with law $\nu = N(0, F'(\sigma) - \overline{F}_\sigma(\sigma))$. 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

$$E_n(g(V_{n,S_n,Q_n})) \to E(g(V)) \tag{73}$$

for all bounded continuous functions g. We have

$$E_n(g(V_{n,S_n,Q_n})) = E_n\left(\int \nu_{n,S_n,Q_n}(dx)g(x)\right).$$

Since $S_n \to \sigma$ and $Q_n \to Q$, we deduce from Lemma 6(b) that $\int \nu_{n,S_n,Q_n}(dx)g(x)$ converges pointwise to $\int \nu(dx)g(x) = E(g(V))$, and that it is bounded by $\|g\|$, so Lebesgue's theorem yields (73). □

LEMMA 8. *The sequence $\widehat{\sigma}_n$ converges in 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 7 implies that $U_{n,S_n,Q_n}(\sigma_n) \to 0$ 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 toward $H(u) := \overline{F}_\sigma(\sigma) - \overline{F}_\sigma(u)$. Hence, $U_{n,S_n,Q_n}(u)$ also converges locally uniformly in u toward $H(u)$, in 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. The definition (68) of $\widehat{\sigma}_n(S_n, Q_n)$ then immediately gives the result. □

PROOF OF THEOREM 6. The proof follows a familiar pattern. As usual, to get the local uniformity in σ for the tightness in (a) [resp. the convergence in (b)], it is enough to obtain the tightness (resp. convergence) under $P_n = P_{n,\sigma_n}$ for any sequence $\sigma_n \to \sigma > 0$. 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. From Lemma 8, we deduce the existence of sets A_n with $P_n(A_n) \to 1$, 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

$$U_n(\sigma_n) = -(\widehat{\sigma}_n - \sigma_n)U_n^{(1)}(T_n) \qquad \text{on the set } A_n. \tag{74}$$

Observe that $U_n^{(1)}(T_n) = -H_{n,S_n}^{(1)}(T_n)$. Since both S_n and T_n converge in probability to σ, (B3) implies that $U_n^{(1)}(T_n) \to -\overline{F}_\sigma^{(1)}(\sigma)$ in probability. Since $\overline{F}_\sigma^{(1)}(\sigma) \neq 0$ by (B4), all of the results of our theorem are now easily deduced from (74) and Lemma 7. □

13. Proofs of the main theorems.

13.1. *Fisher information and LAN.* For Theorems 1 and 2, the key role is played by the density $p_\Delta(\cdot|\sigma, G)$ of the variable X_Δ in (1). From independence of W and Y, we have

$$p_\Delta(x|\sigma, G) = \frac{1}{\sigma\Delta^{1/\beta}} \int G_\Delta(dy) h_\beta\left(\frac{x-y}{\sigma\Delta^{1/\beta}}\right). \tag{75}$$

Since h_β is C^∞ and satisfies (7), it follows that $\sigma \mapsto p_\Delta(x|\sigma, G)$ is also C^∞ and the first two derivatives are

$$\dot{p}_\Delta(x|\sigma, G) = -\frac{1}{\sigma^2 \Delta^{1/\beta}} \int G_\Delta(dy) \breve{h}_\beta\left(\frac{x-y}{\sigma \Delta^{1/\beta}}\right), \tag{76}$$

$$\ddot{p}_\Delta(x|\sigma, G) = \frac{1}{\sigma^3 \Delta^{1/\beta}} \int G_\Delta(dy) \widehat{h}_\beta\left(\frac{x-y}{\sigma \Delta^{1/\beta}}\right), \tag{77}$$

where $\widehat{h}_\beta(x) = h_\beta(x) + 3x h_\beta^{(1)}(x) + x^2 h_\beta^{(2)}(x)$. The Fisher information $I_\Delta(\sigma, G)$ is then

$$I_\Delta(\sigma, G) = \int \frac{\dot{p}_\Delta(x|\sigma, G)^2}{p_\Delta(x|\sigma, G)} \, dx. \tag{78}$$

Taking advantage of (75) and (76), using $b'(G, \alpha)$ and $G'_{\Delta,\alpha}$ [see (13)] and performing the change of variable $x \leftrightarrow (x - \Delta b'(G, \alpha))/\sigma \Delta^{1/\beta}$ in (78) and (80), we obtain, for any $\alpha \in [0, \beta]$,

$$I_\Delta(\sigma, G) = \frac{1}{\sigma^2} \int s_{\Delta,\sigma,G}(x) \, dx \tag{79}$$

$$\text{with } s_{\Delta,\sigma,G}(x) = \frac{(\int G'_{\Delta,\alpha}(du) \breve{h}_\beta(x - u/\sigma))^2}{\int G'_{\Delta,\alpha}(du) h_\beta(x - u/\sigma)}.$$

To prove the LAN property, we also need the Hellinger integral of order $\gamma \in (0, 1)$ between the laws of X_Δ for two different values of σ and the same G, that is,

$$H_\Delta(\gamma \mid \sigma, \sigma', G) = \int p_\Delta(x|\sigma, G)^{1-\gamma} p_\Delta(x|\sigma', G)^\gamma \, dx. \tag{80}$$

The LAN property in Theorem 1 is equivalent to the weak convergence of statistical experiments having log-likelihoods $Z_n(\sigma + u/\sqrt{n}|\sigma, G)$ to a Gaussian shift experiment with unit variance $\mathcal{I}(\beta)/\sigma^2$ (see Definition 80.1, Theorem 80.2 and Corollary 80.6 in [20]). Due to the form of the Hellinger processes for a Gaussian shift and also for i.i.d. observations (see, e.g., [12]), and to Theorem 5.3 of [11], to prove the LAN property, it is enough to prove that for all $u, v \in \mathbb{R}$ and $\gamma \in (0, 1)$,

$$n\big(1 - H_{\Delta_n}(\gamma \mid \sigma + u/\sqrt{n}, \sigma + v/\sqrt{n}, G)\big) \to \frac{\gamma(1-\gamma)(u-v)^2}{2\sigma^2} \mathcal{I}(\beta). \tag{81}$$

PROOF OF (10). Apply the Cauchy–Schwarz inequality with the product of $\breve{h}_\beta/\sqrt{h_\beta}$ and $\sqrt{h_\beta}$ to get

$$s_{\Delta,\sigma,G}(x) \le \int G'_{\Delta,\alpha}(du) \widetilde{h}_\beta(x - u/\sigma).$$

Therefore, (79) yields

$$I_\Delta(\sigma, G) \le \frac{1}{\sigma^2} \int dx \int G'_{\Delta,\alpha}(du) \widetilde{h}_\beta(x - u/\sigma) = \frac{1}{\sigma^2} \mathcal{I}(\beta)$$

[recall (9)] and we have (10). □

PROOF OF THEOREM 2(a). It suffices to prove that if $\sigma_n \to \sigma > 0$, $\Delta_n \to 0$ and G^n is a sequence of measures in $\mathcal{G}(\phi, \alpha)$ for some $\phi \in \Phi$ and some $\alpha \in (0, \beta]$, then

$$I_{\Delta_n}(\sigma_n, G^n) \to \frac{1}{\sigma^2}\mathcal{I}(\beta). \tag{82}$$

Since h_β and $\check{h}_\beta$ are continuous and bounded, Lemma 2(b) yields that $s_{\Delta_n,\sigma_n,G^n} \to \widetilde{h}_\beta$ pointwise. Fatou's lemma yields $\liminf_n I_{\Delta_n}(\sigma_n, G^n) \geq \mathcal{I}(\beta)\sigma^2$. Combining this with (10), we obtain (82). □

PROOF OF THEOREM 2(b). Let $\rho_n = \Delta_n^{1/\alpha_n - 1/\beta}$, which, by our assumption on Δ_n, converges to 1. The measure $G'^n_{\Delta_n,\alpha_n}$ associated with G^n has the density $x \mapsto g_n(x) = h_{\alpha_n}(x\rho_n)/\rho_n$, which converges to $h_\beta(x)$; so, $G'^n_{\Delta_n,\alpha_n,\beta}$ weakly converges to the law with density h_β and, exactly as in the previous proof,

$$s_{\Delta_n,\sigma,G^n}(x) \to s(x) := \frac{(\int h_\beta(u)\check{h}_\beta(x - u\theta/\sigma)\,du)^2}{\int h_\beta(u)h_\beta(x - u\theta/\sigma)\,du}. \tag{83}$$

On the other hand, $|\widetilde{h}_\beta(y)| \leq C(1 \wedge 1/|y|^{1+\beta})$ and $g'_n(y) \leq C(1 \wedge 1/|y|^{1+\alpha_n})$. Using once more the Cauchy–Schwarz inequality, we deduce from (79) that

$$\begin{aligned} s_{\Delta_n,\sigma,G^n}(x) &\leq \int g_n(u)\widetilde{h}_\beta(x - u\theta/\sigma)\,du \\ &\leq s'(x) = C\int\left(1 \wedge \frac{1}{|u|^{1+\beta-\varepsilon}}\right)\left(1 \wedge \frac{1}{|x - u\theta/\sigma|^{1+\beta}}\right)du \end{aligned}$$

for yet another constant C, provided $\alpha_n > \beta - \varepsilon$ for some fixed $\varepsilon \in (0, \beta)$. But, $\int s'(x)\,dx < \infty$, so (83) and the dominated convergence theorem yield

$$I_{\Delta_n}(\sigma, \theta, G^n) \to \frac{1}{\sigma^2}\int s(x)\,dx. \tag{84}$$

Finally, exactly as above, we deduce from the Cauchy–Schwarz inequality and from the fact that the functions $\sqrt{h_\beta}$ and $\check{h}_\beta/\sqrt{h_\beta}$ are not Lebesgue-almost surely multiples of one another, while $h_\beta > 0$ identically, that, in fact, $s(x) < \int h_\beta(u)\widetilde{h}_\beta(x - u\theta/\sigma)\,du$ for all x. Therefore,

$$\begin{aligned} \int s(x)\,dx &< \int dx \int h_\beta(u)\widetilde{h}_\beta(x - u\theta/\sigma)\,du \\ &= \int h_\beta(u)\,du \int \widetilde{h}_\beta(y)\,dy = \int \widetilde{h}_\beta(y)\,dy = \mathcal{I}(\beta) \end{aligned}$$

and (84) yields that $I_{\Delta_n}(\sigma, G^n)$ converges to a limit strictly less that $\mathcal{I}(\beta)/\sigma^2$. □

PROOF OF THEOREM 1. We have proved (10) above and (11) follows from Theorem 2, so it remains to prove LAN, that is, (81). Fix $\sigma > 0$, $G \in \mathcal{G}_\beta$, $\gamma \in (0, 1)$ and $u, v \in \mathbb{R}$. To simplify notation, we write $p_n(x) = p_{\Delta_n}(x|\sigma, G)$, $\dot{p}_n(x) = \dot{p}_{\Delta_n}(x|\sigma, G)$ and $q_n(z, x) = (p_{\Delta_n}(x|\sigma + z/\sqrt{n}, G) - p_n(x))/p_n(x)$. We have

$$H_{\Delta_n}(\gamma \mid \sigma+u/\sqrt{n}, \sigma+v/\sqrt{n}, G) = \int p_n(x)\big(1+q_n(u,x)\big)^\gamma \big(1+q_n(v,x)\big)^{1-\gamma}\,dx.$$

Note that $q_n(z, x) \geq -1$. By a Taylor expansion, we get, for $\delta \in (0, 1)$,

$$\left|\big(1+q_n(z,x)\big)^\delta - 1 - \delta q_n(z,x) + \frac{\delta(1-\delta)}{2} q_n(z,x)^2\right| \leq C|q_n(z,x)|^3 \tag{85}$$

when $|q_n(z, x)| \leq 1/2$ and (85) is trivial when $q_n(z, x) \in [-1, -1/2) \cup (1/2, +\infty)$. Therefore,

$$\begin{aligned}&\Big|\big(1+q_n(u,x)\big)^\gamma\big(1+q_n(v,x)\big)^{1-\gamma} - 1 - \gamma q_n(u,x) - (1-\gamma)q_n(v,x)\\ &\qquad\qquad + \frac{\gamma(1-\gamma)}{2}\big(q_n(u,x)-q_n(v,x)\big)^2\Big|\\ &\quad \leq C\big(|q_n(u,x)|^3 + |q_n(v,x)|^3\big).\end{aligned} \tag{86}$$

Another Taylor expansion gives

$$q_n(z,x) = \frac{z}{\sqrt{n}}\frac{\dot{p}_n(x)}{p_n(x)} + \frac{1}{p_n(x)}\int_0^{z/\sqrt{n}} (z/\sqrt{n} - w)\ddot{p}_{\Delta_n}(x|\sigma+w, G)\,dw. \tag{87}$$

Next, for any $\varepsilon \in (0, 1/2)$, we have $|(y + y')^2 - y^2| \leq 2\varepsilon y^2 + y'^2/\varepsilon$. With $y = (u - v)\dot{p}_n(x)/\sqrt{n}p_n(x)$ and $y' = q_n(u, x) - q_n(v, x) - y$, we deduce

$$\begin{aligned}&\left|\big(q_n(u,x)-q_n(v,x)\big)^2 - \frac{(u-v)^2}{n}\frac{\dot{p}_n(x)^2}{p_n(x)^2}\right|\\ &\quad\leq \frac{C\varepsilon}{n}\frac{\dot{p}_n(x)^2}{p_n(x)^2} + \frac{1}{n\varepsilon p_n(x)^2}\left(\int_{v/\sqrt{n}}^{u/\sqrt{n}} |\ddot{p}_{\Delta_n}(x|\sigma+w,G)|dw\right)^2\end{aligned} \tag{88}$$

(recall that the constant C can change from line to line) and therefore

$$\begin{aligned}&\left|\int p_n(x)\big(q_n(u,x)-q_n(v,x)\big)^2\,dx - \frac{(u-v)^2}{n}I_{\Delta_n}(\sigma,G)\right|\\ &\qquad\leq \frac{C\varepsilon}{n}I_{\Delta_n}(\sigma,G) + \frac{Cb_n}{n\varepsilon},\end{aligned} \tag{89}$$

where

$$b_n = \int \frac{1}{p_n(x)}\left(\int_{v/\sqrt{n}}^{u/\sqrt{n}} |\ddot{p}_{\Delta_n}(x|\sigma+w,G)|dw\right)^2 dx.$$

Observing that $\int p_n(x)\,dx = 1$ and $\int p_n(x)q_n(z, x)\,dx = 0$, (85) and (89) yield

$$\begin{aligned}&\big|1 - H_{\Delta_n}(\gamma \mid \sigma+u/\sqrt{n}, \sigma+v/\sqrt{n}, G) - (u-v)^2\gamma(1-\gamma)I_{\Delta_n}(\sigma,G)/2\big|\\ &\quad\leq \frac{Ca_n}{n} + \frac{C\varepsilon}{n}I_{\Delta_n}(\sigma,G) + \frac{Cb_n}{n\varepsilon},\end{aligned} \tag{90}$$

where, with the notation $z_n = (|u| + |v|)/\sqrt{n}$,

$$a_n = n \int \frac{1}{p_n(x)^2} \biggl(\int_{-z_n}^{z_n} |\dot{p}_{\Delta_n}(x|\sigma + w, G)|\, dw \biggr)^3 dx.$$

Since (90) holds for any $\varepsilon \in (0, 1/2)$ and $I_{\Delta_n}(\sigma, G) \to \mathcal{I}(\beta)/\sigma^2$, we see that (81) holds, provided that we prove $a_n \to 0$ and $b_n \to 0$. Using (75), (76) and (77), we see that if we define

$$c_n = \frac{n^j}{\Delta_n^{1/\beta}} \int \frac{(\int_{-z_n}^{z_n} dw \int G_{\Delta_n}(dy) g(\frac{x-y}{(\sigma+w)\Delta_n^{1/\beta}}))^{2+j}}{(\int G_{\Delta_n}(dy) h_\beta(\frac{x-y}{\sigma \Delta_n^{1/\beta}}))^{1+j}}\, dx,$$

then $a_n \le C c_n$ if $j = 1$ and $g = |\breve{h}_\beta|$, and $b_n \le C c_n$ if $j = 0$ and $g = |\widehat{h}_\beta|$. At this point, we use Hölder's inequality, with conjugate exponents $r = 2 + j$ and $s = (2 + j)/(1 + j)$, first for the integral with respect to dw,

$$\begin{aligned} &\biggl(\int_{-z_n}^{z_n} dw \int G_{\Delta_n}(dy) g\biggl(\frac{x-y}{(\sigma + w)\Delta_n^{1/\beta}} \biggr) \biggr)^{2+j} \\ &\qquad \le \biggl(\int_{-z_n}^{z_n} dw \biggr)^{1+j} \int_{-z_n}^{z_n} \biggl(\int G_{\Delta_n}(dy) g\biggl(\frac{x-y}{(\sigma + w)\Delta_n^{1/\beta}} \biggr) \biggr)^{2+j} dw, \end{aligned}$$

and second for the inside integral with respect to $G_{\Delta_n}(dy)$, to obtain

$$\begin{aligned} &\biggl(\int G_{\Delta_n}(dy) g\biggl(\frac{x-y}{(\sigma + w)\Delta_n^{1/\beta}} \biggr) \biggr)^{2+j} \\ &\qquad \le \biggl(\int h_\beta\biggl(\frac{x-y}{\sigma \Delta_n^{1/\beta}} \biggr) G_{\Delta_n}(dy) \biggr)^{1+j} \\ &\qquad\quad \times \int G_{\Delta_n}(dy) g\biggl(\frac{x-y}{(\sigma + w)\Delta_n^{1/\beta}} \biggr)^{2+j} h_\beta\biggl(\frac{x-y}{\sigma \Delta_n^{1/\beta}} \biggr)^{-(1+j)}. \end{aligned}$$

Therefore,

$$c_n \le \frac{n^j (2z_n)^{1+j}}{\Delta_n^{1/\beta}} \int G_{\Delta_n}(dy) \int_{-z_n}^{z_n} dw \int dx \frac{g(\frac{x-y}{(\sigma+w)\Delta_n^{1/\beta}})^{2+j}}{h_\beta(\frac{x-y}{\sigma \Delta_n^{1/\beta}})^{1+j}}. \tag{91}$$

Next, we introduce the change of variable from x to $z = (x - y)/\Delta_n^{1/\beta}$ to obtain

$$c_n \le n^j (2z_n)^{1+j} \int G_{\Delta_n}(dy) \int_{-z_n}^{z_n} dw \int dz \frac{g(\frac{z}{\sigma+w})^{2+j}}{h_\beta(\frac{z}{\sigma})^{1+j}} \le C n^j z_n^{2+j}.$$

To obtain the last inequality, we note that $\int G_{\Delta_n}(dy) = 1$ and use the fact that $g(\frac{z}{\sigma+w})^{2+j}/h_\beta(\frac{z}{\sigma})^{1+j}$ is less than $C/|z|^{1+\beta}$ for all $|w| \le \sigma/2$, in light of (7). Since $n^j z_n^{2+j} \to 0$ when $j = 0$ and $j = 1$, we obtain $c_n \to 0$ and the proof is finished. $\square$

13.2. *Proof of Theorem 3.* We begin with (b). With the notation $H = \breve{h}_\beta / h_\beta$, we see that in addition to (26), we have $I(k) = E(k(W_1)H(W_1))$ and $\mathcal{I}(\beta) = E(H(W_1)^2)$. 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 $I(k)^2 \le J(k)\mathcal{I}(\beta)$, follows from Cauchy–Schwarz. If $k = \overline{h}_\beta$, we also have $k = 1 + H$, so this inequality is obviously an equality.

For (a), it is enough to prove that if G^n is a sequence in $\mathcal{G}(\phi, \beta)$ for some $\phi \in \Phi$, then the sequence $\sqrt{n}(\widehat{\sigma}_n(G^n, \phi, k) - \sigma)$ converges in law to $N(0, \sigma^2\Sigma^2(k))$, under P_{σ,G^n} and uniformly in σ in compact subsets of $(0, \infty)$. For this, and since $p_n \sim n$, we apply Theorem 6(b) with $P_{n,\sigma} = P_{\sigma,G^n}$ and the p_n increments $\{\chi_i^n\}_{m_n+1 \le i \le n}$, which are independent of the increments $\{\chi_i^n\}_{1 \le i \le m_n}$ and therefore of S_n. The first step consists in proving (A1) for $S_n = S_n(G)$. This amounts to the following lemma, where $\sigma_n \to \sigma > 0$ and $P_n = P_{\sigma_n, G^n}$:

LEMMA 9. *The sequence S_n converges to σ in P_n-probability.*

PROOF. By Lemma 2(b), the variables $Z_{\Delta_n}^n(\beta)$ associated with the law G^n converge in law to 0. The variables $\chi_i'^n$, which equal $\sigma_n W_1 + Z_{\Delta_n}^n(\beta)$ in law, converge in law to σW_1. Hence, $\gamma_n := P_n(|\chi_i'^n| > 1) \to \psi(\sigma)$. If $\zeta_i^n = 1_{\{|\chi_i'^n| > 1\}}$, then (70) applied with $\kappa_n = m_n$ yields $V_n \xrightarrow{P_n} \psi(\sigma)$. Since ψ^{-1} is C^∞ and strictly monotone, the result readily follows. □

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

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

$$H_{n,s}(u) = \Psi_{G^n,\Delta_n,\beta,k_n}\left(\frac{u}{s}, \frac{1}{s}, 0\right).$$

On comparing (24) and (25) with (67) and (68), we see that $\widehat{\sigma}_n(G^n, \phi, k) = \widehat{\sigma}_n(S_n, Q_n)$. Therefore, it remains to prove (B1)–(B6), with a sequence w_n in (B6) satisfying $w_n/\sqrt{p_n} \to \infty$, and that

$$\Xi^2(\sigma) = \sigma^2 J(k)/I(k)^2. \tag{92}$$

Observe that, under P_{σ,G^n}, the variables χ_i^n have the same law as $\sigma W_1 + Z_{\Delta_n}(\beta)$. Then (69) gives $F_{n,s,q}(\sigma) = H_{n,s}(\sigma)$, so (B6) holds with w_n arbitrarily large, while (B2) follows from Lemma 4. If k is bounded, then we have $\|f_{n,s}\| \le \|k\|$ and (B1) is obvious; further, (61) with $\alpha = \beta$ and k^r yields

$$\begin{aligned} &j = 0, 1,\ r = 1, 2,\ \eta \le u \le \frac{1}{\eta},\ v \le \frac{1}{\eta} \\ &\qquad \Longrightarrow \quad \left|\frac{\partial^j}{\partial u^j}\Psi_{G^n,\Delta_n,\beta,k^r}(u, v, 0) - \frac{\partial^j}{\partial u^j}\Psi_{k^r}(u, 0)\right| \le C_{\eta,k}\phi_\beta(\Delta_n^{1/\beta}), \end{aligned}$$

which gives (B3) with $\overline{F}_s(u) = \Psi_k(u/s, 0)$ and (B5) with $F'(u) = \Psi_{k^2}(1, 0)$. On the other hand, when k is unbounded, we have $\|f_{n,s}\| \le \nu_n$ and thus (B1) follows from (23).

Further, $\nu_n \to \infty$ and we can combine (61) for the difference $\frac{\partial^j}{\partial u^j}\Psi_{G^n,\Delta_n,\beta,k_n^r}(u, v, 0) - \frac{\partial^j}{\partial u^j}\Psi_{k_n^r}(u, 0)$, with (54) for the difference $\frac{\partial^j}{\partial u^j}\Psi_{k_n^r}(u, 0) - \frac{\partial^j}{\partial u^j}\Psi_{k^r}(u, 0)$, to obtain, for all n sufficiently large and $j = 0, 1$, $r = 1, 2$, $\eta \le u \le 1/\eta$, $v \le 1/\eta$,

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

Then, in view of (23) and the fact that $2\gamma < \beta$ when $\beta < 2$, we again deduce (B3) with $\bar{F}_s(u) = \Psi_k(u/s, 0)$ and (B5) with $F'(u) = \Psi_{k^2}(1, 0)$.

Since $h_{0,1} = -\breve{h}_\beta$, we deduce that $\overline{F}_\sigma^{(1)}(\sigma) = (1/\sigma)\partial\Psi_k(1, 0)\partial u = -I(k)/\sigma$ [recall (59) and the second part of (20)], hence (B4) holds. We also have $\overline{F}_\sigma(\sigma) = \Psi_k(1, 0) = E(k(W_1))$ and $F'(\sigma) = E(k(W_1)^2)$, hence $J(k) = F'(\sigma) - \overline{F}_\sigma(\sigma)^2$ and (92) follows.

13.3. *Proof of Theorems* 4 *and* 5. As above, we refer to Theorems 4 and 5 as the symmetrical and the asymmetrical cases, respectively. We fix $\alpha \in (0, \beta)$, $\phi \in \overline{\Phi}$ and let $\zeta = \phi_\alpha(1)$ and

$$(93)\qquad \rho = \begin{cases} \rho'(\alpha, \beta), & \text{in the symmetrical case,} \\ \rho(\alpha, \beta), & \text{in the asymmetrical case,} \end{cases} \qquad \lambda_n = \sqrt{n} \wedge \frac{1}{\Delta_n^\rho}.$$

It is enough to take a sequence $\sigma_n \to \sigma > 0$ and a sequence G^n in $\mathcal{G}'(\phi, \alpha)$ in the symmetrical case [resp. $\mathcal{G}(\phi, \alpha)$ in the asymmetrical case] and to prove the tightness or convergence in law of the suitably normalized estimation errors $\widehat{\sigma}_n - \sigma_n$ under the measures $P_n = P_{\sigma_n, G^n}$. Below, we fix the sequences σ_n and G^n and $\Delta_n \to 0$ (so we can assume that $\Delta_n \le 1$ for all n). Let $Z_n := Z^n_{\Delta_n}(\alpha)$ be the variable associated with the measure G^n by (13) and set $b'_n = \Delta_n^{1-1/\beta} b'(G^n, \alpha)$, which vanishes in the symmetrical case.

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 the symmetrical case because $Q_n = 0$. So, we suppose that we are in the asymmetrical case. We introduce some notation. With $j = 1, 2$ and $[\theta^j]^{(p)}$ being the pth derivative of θ^j, let

$$\Gamma_{j,p}(\sigma, u) = (-1)^p E\big([\theta^j]^{(p)}(\sigma W_1 - u)\big) = (-1)^p \int [\theta^j]^{(p)}(\sigma x - u) h_\beta(x)\, dx.$$

Note that $\Gamma_{j,p}(\sigma,u) = \partial^p \Gamma_{j,0}(\sigma,u)/\partial u^p$. Observe that B'_n is the only root of $\mathcal{R}_n(.) = 0$, where

$$\mathcal{R}_n(u) = R_n\big(\Delta_n^{1/\beta}(u + b'_n)\big) = \frac{1}{r_n}\sum_{i=1}^{r_n} \zeta_i^n(u),$$

$$\text{with } \zeta_i^n(u) = \theta(\Delta_n^{-1/\beta}\chi_i^n - u - b'_n).$$

The $\zeta_i^n(u)$'s for $i \geq 1$ are i.i.d. with the same law (under P_n) as the variable $\theta(\sigma_n W_1 + Z_n - u)$. Here, we have used 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 (along with their derivatives), uniformly in u and n, and we can interchange differentiation and expectation. So, we can apply the first part of (51) 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 obtain, for $p, j \in N$,

$$\left|\frac{\partial^p}{\partial u^p}\gamma_{n,j}(u) - \Gamma_{j,p}(\sigma_n,u)\right| \leq C_{p,j}\zeta\Delta_n^\rho. \tag{94}$$

Now, $\mathcal{R}_n$ is also C^∞ and bounded (along with all of its derivatives) uniformly in n, u and ω. Hence, an application of Lemma 5 and the continuity of the functions $\Gamma_{j,p}$ readily yield

$$\frac{\partial^p}{\partial u^p}\mathcal{R}_n(u) \to \Gamma_{1,p}(\sigma,u), \qquad \text{locally uniformly in } u, \text{ in } P_n\text{-probability}, \tag{95}$$

$$\eta_n := \sqrt{r_n}\big(\mathcal{R}_n(0) - \gamma_{n,1}(0)\big) \xrightarrow{L(P_n)} N\big(0, \Gamma_{2,0}(\sigma,0) - \Gamma_{1,0}(\sigma,0)^2\big). \tag{96}$$

The properties of θ imply that $u \mapsto \Gamma_{1,0}(\sigma,.)$ decreases strictly and vanishes at 0 because the function θ is odd. By construction, $\mathcal{R}_n(B'_n) = 0$, so (95), for $p = 0$, implies $B'_n \xrightarrow{P_n} 0$ and also implies $\mathcal{R}_n^{(1)}(B''_n) \xrightarrow{P_n} \Gamma_{1,1}(\sigma,0)$ for any sequence B''_n of random variables converging to 0 in P_n-probability. Since $\mathcal{R}_n(B'_n) = 0$, we have

$$\mathcal{R}_n^{(1)}(B''_n)B'_n = -\mathcal{R}_n(0) = -\frac{\eta_n}{\sqrt{r_n}} - \gamma_{n,1}(0) \tag{97}$$

for some random variable B''_n satisfying $|B''_n| \leq |B'_n|$. Moreover, from $\Gamma_{1,0}(\sigma,0) = 0$, we have $|\gamma_{n,1}(0)| \leq C\zeta\Delta_n^\rho$, by (94). Since $\mathcal{R}_n^{(1)}(B''_n) \xrightarrow{P_n} \Gamma_{1,1}(\sigma,0) \neq 0$, we deduce that $Q_n = \lambda_n B'_n$ satisfies (A2), from (96) [recall $r_n \sim \delta n$ and (93)].

Next, we proceed to prove the consistency of the preliminary estimators S_n. In the symmetrical case, the variables V_n and S_n are the variables $V_n(G^n)$ and $S_n(G^n)$ of (14) and (19), respectively (they do not depend on G^n, in fact), so the result follows from Lemma 9. In the asymmetrical case, set

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

Then (70) yields $V_n(v_n)-\delta_n(v_n)\xrightarrow{P_n}0$ for any sequence v_n. 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 Lemma 2, converges in law to σW_1 provided that $b'_n-\Delta_n^{-1/\beta}v_n\to 0$. Since B_n and $(V_n(v):v\in\mathbb{R})$ are independent and since $B'_n=\Delta_n^{-1/\beta}B_n-b'_n\xrightarrow{P_n}0$ [because $Q_n=\lambda_n B'_n$ satisfies (A2) and $\lambda_n\to\infty$], we deduce that $V_n=V_n(B_n)\xrightarrow{P_n}\psi(\sigma)$. The consistency is then proved as in the end of the proof of Lemma 9.

At this stage, we apply Theorem 6, with the variables (S_n,Q_n) as above and the i.i.d. variables $(\chi^n_{q_n+m_n+i}:1\le i\le p_n)$, which we recall are independent of (S_n,Q_n). With the notation (68) and (31), 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

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

Then (69) gives, for $r=1,2$,

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

Let us check (B1)–(B6). Since k is bounded, (B1) is obvious, whereas (B2) follows from Lemma 3. Next, if we set $\overline{F}_s(u)=\Psi_k(u/s,0)$ and $F'(u)=\Psi_{k^2}(1,0)$, Lemma 4 yields, for $j=0,1$, $\eta\in(0,1)$, $s,u\in[\eta,1/\eta]$ and $|q|\le 1/\eta$,

$$\left|\frac{\partial^j}{\partial u^j}H_{n,s}(u)-\frac{\partial^j}{\partial u^j}\overline{F}_s(u)\right|\le C_{k,\eta}\zeta\Delta_n^\rho,$$
$$|F'_{n,s,q}(u)-F'_s(u)|\le C_{k,\eta}\left(\zeta\Delta_n^\rho+\frac{1}{\lambda_n}\right),$$
$$|F_{n,s,q}(u)-H_{n,s}(u)|\le C_{k,\eta}\left(\zeta\Delta_n^\rho+\frac{1}{\lambda_n}\right).$$

These give (B3) and (B5), as well as (B6) with $w_n=\lambda_n$. Finally, (B4) holds because $\overline{F}_s^{(1)}(s)=(\partial/\partial u)\psi_k(1,0)/s=-I(k)/s$ and (92) holds here as well as in the previous section.

We can thus apply Theorem 6. The sequence $\lambda_n(\widehat{\sigma}_n-\sigma_n)$ is tight under P_n in all cases and this gives the two claims (b). Under (33) or (34), we have $\lambda_n/\sqrt{n}\to\infty$, hence, $\lambda_n\sqrt{p_n}\to\infty$ as well, so $\sqrt{p_n}(\widehat{\sigma}_n-\sigma_n)$ converges in law under P_n to a centered Gaussian variable with variance $\Xi^2(\sigma)=(F'(\sigma)-\overline{F}_\sigma(\sigma)^2)/\overline{F}_\sigma(\sigma)^2$, which, in view of the fact that $\overline{F}_\sigma(\sigma)^2=J(k)/\sigma^2$, equals $\sigma^2\Sigma^2(k)$. Since $p_n\sim n$ in the symmetrical case and $p_n\sim(1-\delta)n$ in the asymmetrical case, we obtain the two claims (a).

Acknowledgments. We are very grateful to two referees, the Associate Editor and Joe Eaton, the Editor, for constructive comments that helped us greatly improve this paper.

REFERENCES

[1] AÏT-SAHALIA, Y. (2004). Disentangling diffusion from jumps. *J. Financial Econ.* **74** 487–528.

[2] AÏT-SAHALIA, Y. and JACOD, J. (2006). Fisher's information for discretely sampled Lévy processes. Technical report, Princeton Univ. and Univ. de Paris 6.

[3] AKGIRAY, V. and LAMOUREUX, C. G. (1989). Estimation of stable-law parameters: A comparative study. *J. Bus. Econom. Statist.* **7** 85–93.

[4] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2006). Impact of jumps on returns and realised variances: Econometric analysis of time-deformed Lévy processes. *J. Econometrics* **131** 217–252. MR2276000

[5] DUMOUCHEL, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimator when sampling from a stable distribution. *Ann. Statist.* **1** 948–957. MR0339376

[6] DUMOUCHEL, W. H. (1973). Stable distributions in statistical inference. I. Symmetric stable distributions compared to other symmetric long-tailed distributions. *J. Amer. Statist. Assoc.* **68** 469–477. MR0378190

[7] DUMOUCHEL, W. H. (1975). Stable distributions in statistical inference. II. Information from stably distributed samples. *J. Amer. Statist. Assoc.* **70** 386–393. MR0378191

[8] FENECH, A. P. (1976). Asymptotically efficient estimation of location for a symmetric stable law. *Ann. Statist.* **4** 1088–1100. MR0426260

[9] FEUERVERGER, A. and MCDUNNOUGH, P. (1981). On efficient inference in symmetric stable laws and processes. In *Statistics and Related Topics* (M. Csörgő, D. A. Dawson, J. N. K. Rao and A. K. Md. Ehsanes Saleh, eds.) 109–122. North-Holland, Amsterdam. MR0665270

[10] FEUERVERGER, A. and MCDUNNOUGH, P. (1981). On the efficiency of empirical characteristic function procedures. *J. Roy. Statist. Soc. Ser. B* **43** 20–27. MR0610372

[11] JACOD, J. (1989). Convergence of filtered statistical models and Hellinger processes. *Stochastic Process. Appl.* **32** 47–68. MR1008908

[12] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. Springer, New York. MR1943877

[13] JEDIDI, W. (2001). Local asymptotic normality of statistical models associated with discrete observations of Lévy processes. Technical report, Johannes Gutenberg Univ. Mainz.

[14] KOUTROUVELIS, I. A. (1980). Regression-type estimation of the parameters of stable laws. *J. Amer. Statist. Assoc.* **75** 918–928. MR0600977

[15] MANCINI, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari* **64** 19–47.

[16] NOLAN, J. P. (1997). Numerical calculation of stable densities and distribution functions. Heavy tails and highly volatile phenomena. *Comm. Statist. Stochastic Models* **13** 759–774. MR1482292

[17] NOLAN, J. P. (2001). Maximum likelihood estimation and diagnostics for stable distributions. In *Lévy Processes: Theory and Applications* (O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick, eds.) 379–400. Birkhäuser, Boston. MR1833706

[18] PRESS, S. J. (1972). Estimation in univariate and multivariate stable distributions. *J. Amer. Statist. Assoc.* **67** 842–846. MR0362666

[19] SINGLETON, K. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *J. Econometrics* **102** 111–141. MR1838137

[20] STRASSER, H. (1985). *Mathematical Theory of Statistics*. *Statistical Experiments and Asymptotic Decision Theory.* de Gruyter, Berlin. MR0812467

[21] WOERNER, J. H. (2004). Power and multipower variation: Inference for high frequency data. Technical report, Univ. Göttingen.

[22] WOERNER, J. H. (2006). Purely discontinuous Lévy processes and power variation: Inference for integrated volatility and the scale parameter. In *Stochastic Finance* (A. N. Shiryaev, M. R. Grosshino, P. Oliveira and M. Esquivel, eds.) 264–276. Springer, Berlin.

[23] ZOLOTAREV, V. M. (1986). *One-Dimensional Stable Distributions*. Amer. Math. Soc., Providence, RI. MR0854867

DEPARTMENT OF ECONOMICS
PRINCETON UNIVERSITY AND NBER
PRINCETON, NEW JERSEY 08544-1021
USA
E-MAIL: yacine@princeton.edu

INSTITUT DE MATHÉMATIQUES DE JUSSIEU
CNRS UMR 7586
UNIVERSITÉ P. ET M. CURIE (PARIS-6)
75252 PARIS CÉDEX 05
FRANCE
E-MAIL: jj@ccr.jussieu.fr