

# JUMP FILTERING AND EFFICIENT DRIFT ESTIMATION FOR LÉVY-DRIVEN SDE'S

BY ARNAUD GLOTER<sup>\*</sup>, DASHA LOUKIANOVA<sup>\*</sup> AND HILMAR MAI<sup>†</sup>

*Université d'Evry Val d'Essonne*<sup>\*</sup> and *ENSAE-ParisTech*<sup>†</sup>

The problem of drift estimation for the solution  $X$  of a stochastic differential equation with Lévy-type jumps is considered under discrete high-frequency observations with a growing observation window. An efficient and asymptotically normal estimator for the drift parameter is constructed under minimal conditions on the jump behavior and the sampling scheme. In the case of a bounded jump measure density these conditions reduce to  $n\Delta_n^{3-\varepsilon} \rightarrow 0$ , where  $n$  is the number of observations and  $\Delta_n$  is the maximal sampling step. This result relaxes the condition  $n\Delta_n^2 \rightarrow 0$  usually required for joint estimation of drift and diffusion coefficient for SDE's with jumps. The main challenge in this estimation problem stems from the appearance of the unobserved continuous part  $X^c$  in the likelihood function. In order to construct the drift estimator we recover this continuous part from discrete observations. More precisely, we estimate, in a non-parametric way, stochastic integrals with respect to  $X^c$ . Convergence results of independent interest are proved for these nonparametric estimators.

**1. Introduction.** The class of solutions of Lévy-driven stochastic differential equations (SDE's) has recently attracted a lot of attention in the literature due to its many applications in various area such as finance, physics and neuroscience. Indeed, it includes important examples taken from finance such as the well-known Barndorff-Nielsen-Shephard model, the Kou model and the Merton model (cf. [2], [13] and [22]) as well as the stochastic Morris-Lecar neuron model (cf. for example [5]) from neuroscience to name just a few. Consequently, statistical inference for these models has recently become an active domain of research.

In this work we aim at estimating the unknown drift parameter  $\theta \in \Theta \subset \mathbb{R}^d$  based on discrete observations  $X_{t_0}^\theta, \dots, X_{t_n}^\theta$  of the process  $X^\theta$  given by

$$(1.1) \quad X_t^\theta = X_0^\theta + \int_0^t b(\theta, X_s^\theta) ds + \int_0^t \sigma(X_s^\theta) dW_s + \int_0^t \gamma(X_{s-}^\theta) dL_s, \quad t \in \mathbb{R}_+,$$

where  $W = (W_t)_{t \geq 0}$  is a one-dimensional Brownian motion and  $L$  a pure jump Lévy process with Lévy measure  $\nu$ .

---

*Keywords and phrases:* Lévy-driven SDE, efficient drift estimation, maximum likelihood estimation, high frequency data, ergodic properties

We consider here the setting of high frequency observations with a growing time window, i.e. for the discrete sample  $X_{t_0}^\theta, \dots, X_{t_n}^\theta$  with  $0 \leq t_0 \leq \dots \leq t_n$  we assume that the sampling step  $\Delta_n := \max\{t_i - t_{i-1} : 1 \leq i \leq n\}$  tends to 0 and  $t_n \rightarrow \infty$  as  $n \rightarrow \infty$ . It is well known that due to the presence of the diffusion part, one can only estimate the drift consistently if  $t_n \rightarrow \infty$ . A crucial point for applications in the high frequency setting is to impose minimal conditions on the sampling step size  $\Delta_n$ . This will be one of our main objectives in this paper.

The topic of high frequency estimation for discretely observed diffusions without jumps is well developed by now. See for example [31], [12] and references therein for joint estimation of drift and diffusion coefficient. Less results are known when a jump component is added to the process. In the case of high frequency estimation for diffusion with an additional jump component, [21] investigates Gaussian quasi-likelihood estimators of a joint drift-diffusion-jump part parameter. A contrast-type estimation function, for joint estimation of drift, diffusion and jump parts when the jumps are of compound Poisson type, is studied in [28]. These results are generalized in [25] to include more general driving Lévy processes, and [23] also provide a large generalization of this paper. The LAN property for drift and diffusion parameters is studied in [29] via Malliavin calculus techniques. In all these papers joint estimation is considered under conditions on the sampling scheme and the Lévy measure, which, in the case of a bounded jump measure density, is at best  $n\Delta_n^2 \rightarrow 0$ .

It is important to note here that the principles of the estimation of the drift, diffusion or jump law parameters are of completely different nature. The estimation of the volatility is feasible on a compact interval, whereas the estimation of the drift and the jump law requires a growing time window. Also due to the Poisson structure of the jump part the estimation of the jump parameter can be well separated from those of the drift and the diffusion part. In this work we focus therefore on the estimation of the drift parameter only and construct a consistent, asymptotically normal and efficient estimator, under conditions on the jump behavior and the sampling scheme, which, in the case of bounded jump measure density reduce to  $n\Delta_n^{3-\varepsilon} \rightarrow 0$ .

A natural approach to estimate the unknown drift parameter would be to use a maximum likelihood estimation, but the likelihood function based on the discrete sample is not tractable in this setting, since it depends on the transition densities of  $X$  which are not explicitly known. On the contrary, the continuous-time likelihood function is explicit. Our aim is to approximate this function from discrete data and hence define some contrast function. The main difficulty is that the continuous-time likelihood involves the continuous part  $X^c$  of  $X$  that is unobservable under discrete sampling. Intuitively, this tells us that the continuous part  $X^c$  has to be recovered, hence the jumps of  $X$  have to be removed in order to obtain an approximation of the continuous likelihood function.

The question of estimation of the continuous part of an Itô-semimartingale appears naturally in many statistical inference questions (cf. for example [3] and [17]) and constitutes in itself an interesting nonparametric problem. In this article we study the question of estimation of stochastic integrals with respect to the continuous part of  $X$  from a discrete

sample of  $X$ . Propositions 4.1 give explicit rates of convergence for our estimators of these quantities. Besides being of independent interest these results constitute the main tool for the asymptotic analysis of our drift estimators.

The technique we use in order to recover stochastic integrals with respect to the continuous part of  $X$  consists in comparing the increments of  $X$  with a threshold, suggested by the typical behavior of a continuous part of  $X$ . This approach will be called jump filtering in the sequel. The thresholding technics is not new and were also used in [28], Mancini [17], Mai [16] and [3]. Note that we do not rise the question of optimality of the threshold's choice, aspect studied for example in Shimizu [26], Shimizu [27] and Figueroa-López and Nisen [6]. In contrast, we have paid particular attention to the improving of the existing conditions on the sampling scheme, which in particular became possible thank to a careful study of the joint law of the biggest jump and of the total contribution of the other jumps in each sampling interval (Lemma 6.3).

The drift estimator is then constructed by applying a jump filter to the discretized likelihood function and maximizing the resulting criterion function to obtain what will be called the filtered MLE (FMLE). To study the properties of the FMLE we first focus on the MLE obtained from continuous observations and show that this MLE is asymptotically normal (Theorem 5.2) with explicit asymptotic variance. We then prove the LAN property which gives by Hájek-Le Cam's convolution theorem that the continuous MLE is efficient (Theorem 5.3). We show in the next step that the FMLE attains asymptotically the same distribution as the MLE based on continuous observations, which proves the efficiency of the FMLE (Theorem 3.2). The last step is mainly based on our results for the jump filter (Proposition 4.1).

The consistency of the FMLE is obtained without further assumptions on the sampling scheme. The asymptotic normality necessitates some additional conditions on the rate at which  $\Delta_n$  goes to 0 that depend on the behavior of the Lévy measure  $\nu$  near zero. In the case where  $\nu$  has a bounded Lebesgue density these conditions reduce to  $n\Delta_n^{3-\varepsilon} \rightarrow 0$  for some  $\varepsilon > 0$ . We believe that this condition is unavoidable, because it is already necessary in the Euler discretization scheme of the stochastic integral with respect to  $X^c$  (Lemma 4.3). This condition appears in [20] for a different parametric drift estimator of stable Levy-driven Ornstein-Uhlenbeck. It is in accordance with the condition  $n\Delta_n^3 \rightarrow 0$  of [7] in the case of drift estimation for continuous diffusions, hence our result can be seen as a generalization of [7] to the presence of jumps.

In the literature on joint estimation of drift and diffusion parameters for models with diffusion and jump part the condition  $n\Delta_n^2 \rightarrow 0$  is usually required (cf. [21], [28] and [25]). The same condition on the sampling scheme appears for joint estimation in the case of continuous diffusions in [31]. Hence, our work shows that by focusing on drift estimation the condition  $n\Delta_n^2 \rightarrow 0$  can be relaxed in the presence of jumps as well.

As will be seen in Section 7 (given in supplemental material [Supplement](#)) many popular models lead to explicit estimators, which do not require the knowledge of the diffusion coefficient and that perform well in numerical examples. In particular, we discuss the practical

choice of the threshold level and its link with the volatility parameter of  $X^c$ .

The structure of the paper is as follows. In Section 2 the problem setting and the main assumptions of this work are introduced. Section 3 contains the construction of the drift estimator from discrete observations together with the main results. In Section 4 we discuss the approximation of the continuous martingale part and state results on the jump filtering error. In Section 5 and 6 we prove the main results and the convergence of the jump filter respectively. The two Sections 7–8 are provided in supplemental material [Supplement](#) of the paper. Section 7 is devoted to applications to popular parametric jump diffusion models and some numerical examples. The section 8 contains the proofs of some auxiliary or technical results used in the paper.

**2. Model, assumptions and ergodicity.** Let  $\Theta$  be a compact subset of  $\mathbb{R}^d$  and  $X^\theta$  a solution to (1.1) which can be rewritten as

$$X_t^\theta = X_0^\theta + \int_0^t b(\theta, X_s^\theta) ds + \int_0^t \sigma(X_s^\theta) dW_s + \int_0^t \int_{\mathbb{R} \setminus \{0\}} \gamma(X_{s-}^\theta) z \mu(ds, dz), \quad t \in \mathbb{R}_+,$$

where  $W = (W_t)_{t \geq 0}$  is a one-dimensional Brownian motion and  $\mu$  is the Poisson random measure on  $[0, \infty) \times \mathbb{R}$  associated with the jumps of the Lévy process  $L = (L_t)_{t \geq 0}$  with Lévy-Khintchine triplet  $(0, 0, \nu)$  such that  $\int_{0 < |z| \leq 1} |z| d\nu(z) < \infty$ . The initial condition  $X_0^\theta$ ,  $W$  and  $L$  are independent.

*2.1. Assumptions.* We suppose that the functions  $b : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  and  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  satisfy the following assumptions:

**ASSUMPTION 1.** *The functions  $\sigma(x), \gamma(x)$  and for all  $\theta \in \Theta$ ,  $b(\theta, x)$  are globally Lipschitz. Moreover, the Lipschitz constant of  $b$  is uniformly bounded on  $\Theta$ .*

Under Assumption 1 equation (1.1) admits a unique non-explosive càdlàg adapted solution possessing the strong Markov property, cf. [1](Theorems 6.2.9. and 6.4.6).

**ASSUMPTION 2.** *For all  $\theta \in \Theta$  there exists a constant  $t > 0$ , such that  $X_t^\theta$  admits a density  $p_t^\theta(x, y)$  with respect to the Lebesgue measure on  $\mathbb{R}$ ; bounded in  $y \in \mathbb{R}$  and in  $x \in K$  for every compact  $K \subset \mathbb{R}$ . Moreover, for every  $x \in \mathbb{R}$ , and every open ball  $U \in \mathbb{R}$  there exists a point  $z = z(x, U) \in \text{supp}(\nu)$  such that  $\gamma(x)z \in U$ .*

The last Assumption was used in [18] to prove the irreducibility of the process  $X^\theta$ . See also [19] for other sets of conditions, sufficient for irreducibility.

- ASSUMPTION 3 (Ergodicity).** (i) *For all  $q > 0$ ,  $\int_{|z| > 1} |z|^q \nu(dz) < \infty$ .*  
(ii) *For all  $\theta \in \Theta$  there exists a constant  $C > 0$  such that  $xb(\theta, x) \leq -C|x|^2$ , if  $|x| \rightarrow \infty$ .*  
(iii)  *$|\gamma(x)|/|x| \rightarrow 0$  as  $|x| \rightarrow \infty$ .*

- (iv)  $|\sigma(x)|/|x| \rightarrow 0$  as  $|x| \rightarrow \infty$ .
- (v)  $\forall \theta \in \Theta, \forall q > 0$  we have  $E|X_0^\theta|^q < \infty$ .

Assumption 2 ensures together with Assumption 3 the existence of unique invariant distribution  $\pi^\theta$ , as well as the ergodicity of the process  $X^\theta$ , as stated in Lemma 2.1 below.

ASSUMPTION 4 (Jumps). (i) *The jump coefficient  $\gamma$  is bounded from below, that is  $\inf_{x \in \mathbb{R}} |\gamma(x)| := \gamma_{min} > 0$  (wlog we suppose  $\gamma_{min} \geq 1$ ).*

*We assume that the set of conditions (ii)–(iv) below holds, or that the condition (ii') below holds:*

- (ii) *the Lévy measure  $\nu$  satisfies  $\int_{0 < |z| \leq 1} |z| \nu(dz) < \infty$ ,*
- (iii) *the Lévy measure  $\nu$  is absolutely continuous with respect to the Lebesgue measure,*
- (iv) *the jump coefficient  $\gamma$  is upper bounded, i.e.  $\sup_{x \in \mathbb{R}} |\gamma(x)| := \gamma_{max} < \infty$ .*
- (ii')  $\nu(\mathbb{R}) < \infty$ .

Note that the integrability condition given by the Assumption 4 (ii) is automatically satisfied in the finite activity case (ii'). The condition (ii) of Assumption 4 insures that the trajectories of the driving Lévy process  $L$  are a.s. of finite variation and hence the integral with respect to  $L$  in (1.1) can be defined as a deterministic Lebesgue-Stieltjes integral. The points (iii), (iv) of the Assumption 4 are technical and needed in the infinite activity case. They can be removed in the simpler case of finite jump activities, and this is why we introduce the two sets of assumptions (ii)–(iv) and (ii').

The following assumption insures the existence of the likelihood function.

ASSUMPTION 5 (Non-degeneracy). *There exists some  $\alpha > 0$ , such that  $\sigma^2(x) \geq \alpha$  for all  $x \in \mathbb{R}$ .*

ASSUMPTION 6 (Identifiability). *For all  $\theta \neq \theta', (\theta, \theta') \in \Theta^2$ ,*

$$\int_{\mathbb{R}} \frac{(b(\theta, x) - b(\theta', x))^2}{\sigma^2(x)} d\pi^\theta(x) > 0.$$

We can see (cf. Proposition 8.1 in Supplement) that this last assumption is equivalent to

$$(2.1) \quad \forall \theta \neq \theta', \quad (\theta, \theta') \in \Theta^2, \quad b(\theta, \cdot) \neq b(\theta', \cdot).$$

For  $f : \Theta \rightarrow \mathbb{R}$  denote by  $\nabla_\theta f : \Theta \rightarrow \mathbb{R}^d$  the gradient column vector and by  $\partial_\theta^2 f := \left( \partial_{\theta_i, \theta_j}^2 f \right)_{1 \leq i, j \leq d}$  the Hessian matrix of  $f$ . We define  $|\theta|$  as the Euclidian norm of  $\theta \in \mathbb{R}^d$ , and  $|\partial_\theta^2 f| := \sqrt{\sum_{i, j=1}^d |\partial_{\theta_i, \theta_j}^2 f|^2}$  as the Euclidian norm of the Hessian matrix of  $f$ . We say that  $f : \mathbb{R} \rightarrow \mathbb{R}$  is sub-polynomial, if there exists some polynomial function  $P : \mathbb{R} \rightarrow \mathbb{R}$  such that  $|f(x)| \leq |P(x)|$  for all  $x \in \mathbb{R}$ . The following assumption is used to insure the uniform in  $\theta$  convergence needed in the proofs of consistency and asymptotic normality:

ASSUMPTION 7 (Hölder-continuity of drift). **(i)** For all  $x \in \mathbb{R}$ ,  $b(\cdot, x)$  is Hölder-continuous with respect to  $\theta \in \Theta$ :

$$\forall \theta, \theta', \quad |b(\theta, x) - b(\theta', x)| \leq K(x)|\theta - \theta'|^\kappa,$$

where  $0 < \kappa \leq 1$  and  $K : \mathbb{R} \rightarrow \mathbb{R}_+$  is sub-polynomial.

**(ii)** For all  $x \in \mathbb{R}$ ,  $b(\cdot, x)$  is twice continuously differentiable with respect to  $\theta$  and  $\nabla b(\cdot, x)$  and  $\partial^2 b(\cdot, x)$  are Hölder-continuous with respect to  $\theta \in \Theta$ :

$$\forall \theta, \theta', \quad |\nabla b(\theta, x) - \nabla b(\theta', x)| \leq K_1(x)|\theta - \theta'|^{\kappa_1}$$

$$\forall \theta, \theta', \quad |\partial_{\theta}^2 b(\theta, x) - \partial_{\theta}^2 b(\theta', x)| \leq K_2(x)|\theta - \theta'|^{\kappa_2}$$

where  $0 < \kappa_1, \kappa_2 \leq 1$  and  $K_1, K_2 : \mathbb{R} \rightarrow \mathbb{R}_+$  are sub-polynomial.

We also need the following technical assumption:

ASSUMPTION 8. The functions  $b, \sigma, \nabla_{\theta} b, \partial_{\theta}^2 b$  are twice continuously differentiable with respect to  $x$ . The functions  $\sigma', \sigma''$  as well as the functions

$$x \mapsto \sup_{\theta \in \Theta} \left| \frac{\partial^{i+j} b(\theta, x)}{\partial^i x \partial^j \theta} \right|$$

are sub-polynomial for all  $0 \leq i \leq 2$  and  $0 \leq j \leq 2$ .

Define the asymptotic Fisher information by

$$(2.2) \quad I(\theta) = \left( \int_{\mathbb{R}} \frac{\partial_{\theta_i} b(\theta, x) \partial_{\theta_j} b(\theta, x)}{\sigma^2(x)} \pi^{\theta}(dx) \right)_{1 \leq i, j \leq d}.$$

ASSUMPTION 9. For all  $\theta \in \Theta$ ,  $I(\theta)$  is non-degenerated.

**2.2. Ergodic properties of solutions.** In all our statistical analysis an important role is played by ergodic properties of solutions of equation (1.1). The following lemma is a generalization of a result of [18]. It states conditions for the existence of an invariant measure  $\pi^{\theta}$  such that an ergodic theorem holds and moments of all order exist. A proof is given in Section 8 of [Supplement](#).

LEMMA 2.1. Under Assumptions 1 to 4, for all  $\theta \in \Theta$ ,  $X^{\theta}$  admits a unique invariant distribution  $\pi^{\theta}$  and the ergodic theorem holds:

(1) for every measurable function  $g : \mathbb{R} \rightarrow \mathbb{R}$  satisfying  $\pi^{\theta}(g) < \infty$ , we have a.s.

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t g(X_s^{\theta}) ds = \pi^{\theta}(g).$$

- (2) For all  $q > 0$ ,  $\pi^\theta(|x|^q) < \infty$ .  
(3) For all  $q > 0$ ,  $\sup_{t \in \mathbb{R}} E[|X_t^\theta|^q] < \infty$  and  $\sup_{t \in \mathbb{R}} E[|X_{t-}^\theta|^q] < \infty$ .  
(4) Moreover,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t E[|X_s^\theta|^q] ds = \pi^\theta(|x|^q).$$

**3. Construction of the estimator and main results.** We define a discrete approximation of the continuous time likelihood function by employing a jump filtering technique and hence obtain an approximate maximum likelihood estimator. We prove that this drift estimator attains asymptotically the same performance as the maximum likelihood estimator based on continuous observations under suitable assumptions on the jump behavior of the driving Lévy process  $L$ .

3.1. *Construction of the estimator.* Let  $X^\theta$  be given by (1.1). We denote by  $P^\theta$  the law of  $X^\theta$  on the Skorokhod space  $D[0, \infty)$  of real-valued càdlàg functions, and  $P_t^\theta$  its restriction on  $D[0, t)$ . From now on we denote the true parameter value by  $\theta^*$ , an interior point of the parameter space  $\Theta$  that we want to estimate. We shorten  $X$  for  $X^{\theta^*}$  and  $P, E, \pi$  for respectively  $P^{\theta^*}, E^{\theta^*}, \pi^{\theta^*}$ . Suppose that we observe a finite sample

$$X_{t_0}, \dots, X_{t_n}; \quad 0 = t_0 \leq t_1 \leq \dots \leq t_n.$$

Every observation time point depends also on  $n$ , but to simplify our notation we suppress this index. We will be working in a high-frequency setting, i.e.

$$\Delta_n := \sup_{i=0, \dots, n-1} (t_{i+1} - t_i) \xrightarrow{n \rightarrow \infty} 0.$$

We assume  $\lim_{n \rightarrow \infty} t_n = \infty$  and  $n\Delta_n = O(t_n)$  as  $n \rightarrow \infty$ . Under Assumption 5,  $P_t^\theta$  and  $P_t$  are mutually locally absolutely continuous for any  $\theta \in \Theta$  (cf. for example [11]). We define the likelihood function by

$$(3.1) \quad \mathcal{L}_t(\theta, X) = \exp \left( \int_0^t \sigma(X_s)^{-2} b(\theta, X_s) dX_s^c - \frac{1}{2} \int_0^t \sigma(X_s)^{-2} b(\theta, X_s)^2 ds \right),$$

and the log-likelihood function as

$$(3.2) \quad \ell_t(\theta) := \ln \mathcal{L}_t(\theta, X).$$

Note that our choice for the likelihood  $\mathcal{L}_t(\theta, X)$  differs from the Radon Nicodym density  $\frac{dP_t^\theta}{dP_t}$  by the multiplicative factor not depending on  $\theta$ . The crucial point here is the appearance of  $X^c$  in (3.1), since when  $X$  is observed discretely, its continuous part remains unknown. To handle this problem we use a jump filter as described below.

For  $g : [0, t_n] \rightarrow \mathbb{R}$ , set  $\Delta_i^n g = g_{t_i} - g_{t_{i-1}}$ ,  $i = 1, \dots, n$ . In particular,  $\Delta_i^n X = X_{t_i} - X_{t_{i-1}}$ ,  $\Delta_i^n X^c = X_{t_i}^c - X_{t_{i-1}}^c$  and  $\Delta_i^n Id = t_i - t_{i-1}$ . Let  $(a_n^i), i = 1, \dots, n$ , be a sequence of positives

random variables, bounded from above and below by some constants  $\underline{a}$ ,  $\bar{a}$ :  $0 < \underline{a} \leq a_n^i \leq \bar{a} < \infty$  and such that  $a_n^i$  is measurable with respect to the observations  $\{X_{t_j}; j < i\}$ . Let  $\varepsilon \in (0, 1/2)$  and denote

$$(3.3) \quad v_n^i = a_n^i v_n, \quad v_n = \Delta_n^{1/2-\varepsilon}, n \geq 1, i = 1 \dots, n.$$

Define a discrete, jump-filtered approximation  $\ell_{t_n}^n$  of the log-likelihood function as follows.

$$(3.4) \quad \ell_{t_n}^n(\theta) = \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} b(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} - \frac{1}{2} \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} b(\theta, X_{t_{i-1}})^2 \Delta_i^n Id.$$

The cut-off sequence  $(v_n^i)$  is chosen in order to asymptotically filter the increments of  $X$  containing jumps. The increments of the continuous martingale part are typically of the order  $\Delta_n^{1/2}$  which leads to the definition (3.3). We allow the threshold level associated to  $\Delta_i^n X$  to depend on a random coefficient  $a_n^i$  based on the past observations  $(X_{t_j})_{j < i}$ . Although the asymptotic results we obtain do not depend on the choice of the  $a_n^i$ , numerical simulations show that choosing properly these weights is crucial when dealing with finite sample. The challenge now is to find suitable conditions on  $\Delta_n$ ,  $\varepsilon$  and  $\nu$  to make the likelihood (3.2) well approximated by its discretized and jump filtered counterpart (3.4) even in the case of infinite activity. Of course, we can choose  $\varepsilon$  arbitrarily small, which is a choice we have in mind. Finally, we define an estimator  $\hat{\theta}_n$  of  $\theta^*$  as

$$(3.5) \quad \hat{\theta}_n \in \operatorname{argmax}_{\theta \in \Theta} \ell_{t_n}^n(\theta)$$

and in the sequel we call it the filtered MLE (FMLE).

**3.2. Main results.** The following theorem gives a general consistency result for the FMLE  $\hat{\theta}_n$  that holds for finite and infinite activity without further assumptions on  $n$ ,  $\Delta_n$  and  $v_n$ .

**THEOREM 3.1 (Consistency).** *Suppose that Assumptions 1 to 8 hold, then the FMLE  $\hat{\theta}_n$  is consistent in probability:*

$$\hat{\theta}_n \xrightarrow{P} \theta^*, \quad n \rightarrow \infty.$$

Recall that the Fisher information  $I$  is given by (2.2). The following theorem gives the asymptotic normality of the FMLE.

**THEOREM 3.2 (Asymptotic normality).** *Suppose that Assumptions 1 to 9 hold. Assume furthermore that  $n\Delta_n^{3-\varepsilon} \rightarrow 0$ ,*

$$\sqrt{n}\Delta_n^{3/2-2\varepsilon} \left( \int_{|z| \geq 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2} \rightarrow 0 \quad \text{and} \quad \sqrt{n}\Delta_n \left( \int_{|z| \leq 3\bar{a}v_n} |z|\nu(dz) \right)^{1-\varepsilon/2} \rightarrow 0$$



as  $n \rightarrow \infty$ . Then the FMLE  $\hat{\theta}_n$  is asymptotically normal:

$$t_n^{1/2}(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} N(0, I^{-1}(\theta^*)), \quad n \rightarrow \infty.$$

Furthermore, the FMLE  $\hat{\theta}_n$  is asymptotically efficient in the sense of the Hájek-Le Cam convolution theorem.

REMARK 1. In the case of finite activity and if  $\nu$  has a bounded Lebesgue density, the conditions relating  $n, \Delta_n$  and  $v_n$  in the Theorem 3.2 reduce to  $n\Delta_n^{3-\tilde{\varepsilon}} \rightarrow 0$ , where  $\tilde{\varepsilon} > 0$  is arbitrarily small.

EXAMPLE 3.3 (generalized tempered stable jumps). To illustrate the influence of the jump behavior of  $L$  on the conditions on  $n$  and  $\Delta_n$  given in Theorem 3.2, let us consider the example of a generalized tempered  $\alpha$ -stable driving Lévy process (cf. [4], p.119). In this case the Lévy density is given by  $\nu(dz) = C|z|^{-(1+\alpha)}e^{-\lambda|z|}dz$ , with  $\lambda > 0$ ,  $C > 0$  and satisfies the conditions of Theorem 3.2 if  $\alpha < 1$ . This density is bounded for  $\alpha \leq -1$ , unbounded and integrable for  $-1 < \alpha < 0$  and unbounded and non-integrable for  $0 \leq \alpha < 1$ .

The conditions on  $n$ ,  $\Delta_n$  and  $\nu$  in Theorem 3.2 can be summarized in  $n\Delta_n^{2-\alpha-\tilde{\varepsilon}} \vee n\Delta_n^{3-4\tilde{\varepsilon}} \rightarrow 0$  for some explicit  $\tilde{\varepsilon} > 0$ . We observe that a higher Blumenthal-Gettoor index  $\alpha$  requires a faster convergence of  $\Delta_n$  to zero. This in line with the intuition that when the intensity of small jumps increases (i.e.  $\alpha$  increases) more and more frequent observations are needed to have a sufficient performance of the jump filter.

**4. Nonparametric estimation of  $X^c$  via jump filtering.** The estimation problem considered in this work leads naturally to the more fundamental problem of approximation of the continuous martingale part  $X^c$  from discrete observations of a jump diffusion  $X$ . In this section we prove approximation results of this sort for integral functionals with respect to  $X^c$ . Since we need both uniform and non-uniform versions for the drift estimation problem, both settings will be discussed. Recall that the cut-off sequence  $v_n^i$  as well as  $v_n$  were defined in (3.3). We denote by  $\xrightarrow{P}$  the convergence in  $P$ -probability. If  $(Z_n)_{n \in \mathbb{N}}$  is a sequence of random variables and  $(u_n)_{n \in \mathbb{N}}$  is a positive valued sequence, we say that  $Z_n = O_{\mathbb{L}^1}(u_n)$  (resp.  $Z_n = o_{\mathbb{L}^1}(u_n)$ ) if  $E|Z_n| = O(u_n)$  (resp.  $E|Z_n| = o(u_n)$ ). We say that  $Z_n = o_P(u_n)$  if  $Z_n/u_n \xrightarrow{P} 0$ .

PROPOSITION 4.1 (jump filtering). Suppose that Assumptions 1 to 4 hold. Suppose that  $f : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$  satisfies:

a) for all  $x \in \mathbb{R}$ ,  $\theta \mapsto f(\theta, x)$  is  $\kappa$ -Hölder continuous for some  $0 < \kappa \leq 1$ :

$$\forall \theta, \theta', \quad |f(\theta, x) - f(\theta', x)| \leq C(x)|\theta - \theta'|^\kappa,$$

where  $C : \mathbb{R} \rightarrow \mathbb{R}_+$  is sub-polynomial;

b) for all  $\theta \in \Theta$ ,  $f(\theta, \cdot) \in \mathcal{C}^2(\mathbb{R})$  and  $\sup_{\theta \in \Theta} |f(\theta, \cdot)|$ ,  $\sup_{\theta \in \Theta} |f'_x(\theta, \cdot)|$  and  $\sup_{\theta \in \Theta} |f''_x(\theta, \cdot)|$  are sub-polynomial.

Then the following statements hold:

(i) without any assumption on the way that  $\Delta_n \rightarrow 0$  as  $n \rightarrow \infty$ ,

$$(n\Delta_n)^{-1} \sup_{\theta \in \Theta} \left| \int_0^{t_n} f(\theta, X_s) dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right| \xrightarrow{P} 0;$$

(ii) if  $n\Delta_n^{3-\varepsilon} \rightarrow 0$ ,

$$\sqrt{n}\Delta_n^{3/2-\varepsilon} \left( \int_{|z| \geq 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2} \rightarrow 0 \quad \text{and} \quad \sqrt{n}\Delta_n \left( \int_{|z| \leq 3\bar{a}v_n} |z|\nu(dz) \right)^{1-\varepsilon/2} \rightarrow 0$$

as  $n \rightarrow \infty$ , then for any  $\theta \in \Theta$ ,

$$(n\Delta_n)^{-1/2} \left| \int_0^{t_n} f(\theta, X_s) dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right| \xrightarrow{P} 0.$$

The proof of Proposition 4.1 is based on the following two lemmas. Lemma 4.2 describes the approximation of the discretized stochastic integral with respect to  $X^c$  by the jump filter. Lemma 4.3 describes the convergence of the Euler scheme in order to approximate the stochastic integral with respect to  $X^c$  by the corresponding discrete sum. All the results of this section are proved in Section 6.

LEMMA 4.2 (jump filtering error). *Assume that  $X$  satisfies Assumption 1 to 4 and  $f : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$  is such that  $\sup_{\theta \in \Theta} |f(\theta, x)|$  is sub-polynomial. Then the following hold:*

(i) without any assumption on the way that  $\Delta_n \rightarrow 0$  as  $n \rightarrow \infty$ ,

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left( \Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right) \right| = \\ & O_{\mathbb{L}^1} \left( n\Delta_n^{3/2-\varepsilon/2} \left( \int_{|z| \geq \underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2} + n\Delta_n \left( \int_{|z| \leq 3\bar{a}v_n} |z|\nu(dz) \right)^{1-\varepsilon/2} + n\Delta_n^2 \right); \end{aligned}$$

(ii) for all  $\theta \in \Theta$ , if  $n\Delta_n^{3-\varepsilon} \left( \int_{|z| \geq 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{2-\varepsilon} \rightarrow 0$ , as  $n \rightarrow \infty$ , then

$$\begin{aligned} & \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left( \Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right) = o_P \left( \sqrt{n\Delta_n} \right) + O_{\mathbb{L}^1} \left( n\Delta_n^2 \right) \\ & + o_{\mathbb{L}^1} \left( n\Delta_n^{2-\varepsilon} \left( \int_{|z| \geq 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2} \right) + O_{\mathbb{L}^1} \left( n\Delta_n \left( \int_{|z| \leq 3\bar{a}v_n} |z|\nu(dz) \right)^{1-\varepsilon/2} \right). \end{aligned}$$

The approximation of the stochastic integral is treated in the following lemma.

LEMMA 4.3 (Euler scheme). *Suppose that  $f : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$  satisfies the following assumptions:*

a) *for all  $x \in \mathbb{R}$ ,  $\theta \mapsto f(\theta, x)$  is  $\kappa$ -Hölder continuous for some  $0 < \kappa \leq 1$ :*

$$\forall \theta, \theta', \quad |f(\theta, x) - f(\theta', x)| \leq K(x)|\theta - \theta'|^\kappa;$$

*where  $K : \mathbb{R} \rightarrow \mathbb{R}_+$  is sub-polynomial;*

b) *for all  $\theta \in \Theta$ ,  $x \mapsto f(\theta, x) \in \mathcal{C}^2(\mathbb{R})$  and  $\sup_{\theta \in \Theta} |f(\theta, \cdot)|$ ,  $\sup_{\theta \in \Theta} |f'_x(\theta, \cdot)|$  and  $\sup_{\theta \in \Theta} |f''_x(\theta, \cdot)|$  are sub-polynomial.*

Under Assumptions 1 to 4, we obtain

(i) *without any assumption on the way that  $\Delta_n \rightarrow 0$ , as  $n \rightarrow \infty$ ,*

$$\sup_{\theta \in \Theta} (n\Delta_n)^{-1} \left| \int_0^{t_n} f(\theta, X_s) dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X^c \right| \xrightarrow{P} 0;$$

(ii) *if  $n\Delta_n^{3-\varepsilon} \rightarrow 0$ , then, as  $n \rightarrow \infty$ ,*

$$\forall \theta \in \Theta, \quad (n\Delta_n)^{-1/2} \left| \int_0^{t_n} f(\theta, X_s) dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X^c \right| \xrightarrow{P} 0.$$

When discretizing the likelihood function, we need the following lemma, whose proof can be found in the Section 6.

LEMMA 4.4. *Suppose that Assumptions 1-4 are satisfied. Suppose that  $f : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$  is such that  $\forall \theta \in \Theta$ ,  $f(\theta, \cdot) \in \mathcal{C}^1(\mathbb{R})$  and  $\sup_{\theta \in \Theta} |f'(\theta, \cdot)|$  is sub-polynomial. Then we obtain:*

(i) *as  $n \rightarrow \infty$ ,*

$$\sup_{\theta \in \Theta} \left| \int_0^{t_n} f(\theta, X_s) ds - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n Id \right| = O_{\mathbb{L}^1}(n\Delta_n^{3/2});$$

(ii) *if  $n\Delta_n^{3-\varepsilon} \rightarrow 0$ , then*

$$\forall \theta \in \Theta, \quad (n\Delta_n)^{-1/2} \left| \int_0^{t_n} f(\theta, X_s) ds - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n Id \right| \xrightarrow{P} 0.$$

## 5. Proofs of main results.

5.1. *MLE for continuous observations.* Let  $\bar{\theta}_t$  be the true MLE maximizing the log-likelihood function given by (3.2) and based on continuous observations :

$$(5.1) \quad \bar{\theta}_t \in \operatorname{argmax}_{\theta \in \Theta} \ell_t(\theta).$$

Before moving to discrete observations we prove here some asymptotic results for  $\bar{\theta}_t$ . This is a first step in order to prove the asymptotic results for the FMLE.

THEOREM 5.1. *Suppose that Assumptions 1–6 and 7(i) are satisfied. Then*

$$\lim_{t \rightarrow \infty} \bar{\theta}_t = \theta^* \quad P - a.s.$$

PROOF. Denote

$$(5.2) \quad \tilde{\ell}_t(\theta) := \int_0^t \frac{(b(\theta, X_s) - b(\theta^*, X_s))}{\sigma(X_s)} dW_s - \frac{1}{2} \int_0^t \frac{(b(\theta, X_s) - b(\theta^*, X_s))^2}{\sigma^2(X_s)} ds.$$

Using (1.1) and the fact that the observed trajectory corresponds to the true value of parameter  $\theta^*$ , we can easily see that

$$(5.3) \quad \ell_t(\theta) = \tilde{\ell}_t(\theta) + \int_0^t \frac{b(\theta^*, X_s)}{\sigma(X_s)} dW_s + \frac{1}{2} \int_0^t \frac{b^2(\theta^*, X_s)}{\sigma^2(X_s)} ds.$$

The difference between  $\ell(\theta)$  and  $\tilde{\ell}_t(\theta)$  does not depend on  $\theta$ , hence also  $\bar{\theta}_t \in \operatorname{argmax}_{\theta \in \Theta} \tilde{\ell}_t(\theta)$ . For  $\theta \in \Theta$ , define  $M_t(\theta) := \int_0^t \frac{(b(\theta, X_s) - b(\theta^*, X_s))}{\sigma(X_s)} dW_s$ . The process  $(M_t(\theta), t \geq 0)$  is a continuous local martingale, with quadratic variation given by  $A_t(\theta) := \langle M(\theta) \rangle_t = \int_0^t \frac{(b(\theta, X_s) - b(\theta^*, X_s))^2}{\sigma^2(X_s)} ds$ . Note that  $\tilde{\ell}_t(\theta) = -\frac{1}{2}A_t(\theta) + M_t(\theta)$ . Recall that  $\pi$ , given by the Lemma 2.1 is an invariant distribution of  $X$  and denote

$$(5.4) \quad \tilde{\ell}(\theta) = -\frac{1}{2}\pi \left( \frac{(b(\theta, \cdot) - b(\theta^*, \cdot))^2}{\sigma^2(\cdot)} \right).$$

Using Assumptions 5, 7(i) and Lemma 2.1(2), we see that for all  $\theta \in \Theta$ ,  $\tilde{\ell}(\theta) \in \mathbb{R}$ . Hence, using the Lemma 2.1(1) for all  $\theta \in \Theta$ ,

$$\lim_{t \rightarrow \infty} -\frac{1}{2t}A_t(\theta) = \tilde{\ell}(\theta) \quad P - a.s.$$

Moreover, using Assumptions 1, 5 and 7(i) we can see that the family

$$(5.5) \quad \left\{ \frac{1}{t}A_t(\theta) \right\}_{t>0} \text{ is equicontinuous } P - a.s.$$

Indeed,  $\frac{1}{t}|A_t(\theta) - A_t(\theta')| \leq C|\theta - \theta'|^\kappa \frac{1}{t} \int_0^t |\tilde{K}(X_s)| ds$ , where  $\tilde{K}$  is some polynomial function. Using ergodic theorem, which holds thanks to the Lemma 2.1,  $\frac{1}{t} \int_0^t |\tilde{K}(X_s)| ds$  converges almost surely to some finite limit. Hence (5.5) follows. As a consequence,

$$(5.6) \quad \limsup_{t \rightarrow \infty} \sup_{\theta \in \Theta} \left| -\frac{1}{2t} A_t(\theta) - \tilde{\ell}(\theta) \right| = 0 \quad P - a.s.$$

Denote  $A_t(\theta, \theta') := \langle M_t(\theta) - M_t(\theta') \rangle_t$ . Using Assumptions 5 and 7(i), for all  $(\theta, \theta') \in \Theta^2$ ,

$$A_t(\theta, \theta') \leq C|\theta - \theta'|^{2\kappa} V_t,$$

where  $V_t := \int_0^t \left( \frac{K^2(X_s)}{\sigma^2(X_s)} \vee 1 \right) ds \rightarrow \infty$ , if  $t \rightarrow \infty$ . Therefore all assumptions of the Theorem 2 in [15] are satisfied. As a conclusion, the family  $\left\{ \frac{M_t(\theta)}{A_t(\theta)}; \theta \in \Theta, t \geq 0 \right\}$  satisfies the Uniform Law of Large Numbers on any compact  $K \in \Theta$  not containing  $\theta^*$ , i.e.

$$\limsup_{t \rightarrow \infty} \sup_{\theta \in K} \left| \frac{M_t(\theta)}{A_t(\theta)} \right| = 0.$$

We deduce, using (5.6), that  $\lim_{t \rightarrow \infty} \sup_{\theta \in K} \left| \frac{M_t(\theta)}{t} \right| = 0$ , and hence,  $P - a.s.$

$$(5.7) \quad \sup_{\theta \in K} |t^{-1} \tilde{\ell}_t(\theta) - \tilde{\ell}(\theta)| \rightarrow 0.$$

We can now derive the a.s. consistency of  $\bar{\theta}_t$  following classical Wald's method. We refer for instance to Theorem 5.7 in [30] for a simple presentation of Wald's approach, and stress out the fact that all convergences and hence consistency holds  $P$ -a.s. in our setting. Indeed, observe that  $\tilde{\ell}(\theta) \leq 0$ , and  $\tilde{\ell}(\theta) = 0 \iff \theta = \theta^*$ , which implies

$$(5.8) \quad \sup_{\theta: d(\theta, \theta^*) \geq \varepsilon} \tilde{\ell}(\theta) < \tilde{\ell}(\theta^*).$$

We deduce from (5.7) and (5.8) that  $P$ -a.s. for all  $\varepsilon > 0$ ,  $\lim_{t \rightarrow \infty} \sup_{d(\theta, \theta^*) \geq \varepsilon} \frac{1}{t} \tilde{\ell}_t(\theta) < \tilde{\ell}(\theta^*)$  and hence for  $t > t(\omega)$  large enough  $\sup_{d(\theta, \theta^*) \geq \varepsilon} \tilde{\ell}_t(\theta) < \tilde{\ell}_t(\theta^*)$ . It yields, for  $t > t(\omega)$ ,  $d(\bar{\theta}_t, \theta^*) < \varepsilon$ , which means the a.s. consistency.  $\square$

Recall that  $I$  is the Fisher information given by (2.2).

The next result is a central limit theorem for the estimation error. It is important for us in the sequel, since the asymptotic variance serves as a benchmark for the case of discrete observations.

**THEOREM 5.2.** *Suppose that Assumptions 1–9 hold. Then the MLE  $\bar{\theta}_t$  is asymptotically normal:*

$$t^{1/2}(\bar{\theta}_t - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(\theta^*)) \quad \text{as } t \rightarrow \infty.$$

PROOF. Due to Assumptions 5 and 7, Theorem 2.2 in [9] and Theorem 1 in [15] for all  $t > 0$  the criterion function  $\tilde{\ell}_t(\theta, X)$  is twice continuously differentiable in  $\theta$ .

From (5.2) the score function can be written as  $\nabla_{\theta} \ell = \nabla_{\theta} \tilde{\ell} = (\partial_{\theta_1} \tilde{\ell}_t, \dots, \partial_{\theta_d} \tilde{\ell}_t)^T$  where

$$(5.9) \quad \partial_{\theta_i} \tilde{\ell}_t(\theta) = - \int_0^t \frac{(b(\theta, X_s) - b(\theta^*, X_s)) \partial_{\theta_i} b(\theta, X)}{\sigma^2(X_s)} ds + \int_0^t \frac{\partial_{\theta_i} b(\theta, X_s)}{\sigma(X_s)} dW_s,$$

for  $i = 1, \dots, d$ . A Taylor expansion around  $\bar{\theta}_t$  yields

$$(5.10) \quad t^{-1} \int_0^1 \partial_{\bar{\theta}}^2 \tilde{\ell}_t(\theta^* + s(\bar{\theta}_t - \theta^*)) ds \times \sqrt{t}(\bar{\theta}_t - \theta^*) = -\frac{1}{\sqrt{t}} \nabla_{\theta} \tilde{\ell}_t(\theta^*).$$

where  $\partial_{\bar{\theta}}^2 \tilde{\ell}_t$  is the Hessian matrix of  $\tilde{\ell}$ . Hence, to obtain a CLT for the estimation error  $t^{1/2}(\bar{\theta}_t - \theta^*)$  we will first show the convergence of the right hand side in (5.10). The equation (5.9) gives for  $\theta = \theta^*$ ,  $\nabla_{\theta} \tilde{\ell}_t(\theta^*) = \int_0^t \frac{\nabla_{\theta} b(\theta^*, X_s)}{\sigma(X_s)} dW_s$ , such that the central limit theorem for multidimensional local martingales [14] implies

$$(5.11) \quad t^{-1/2} \nabla_{\theta} \tilde{\ell}_t(\theta^*) = t^{-1/2} \int_0^t \frac{\nabla_{\theta} b(\theta^*, X)}{\sigma(X_s)} dW_s \xrightarrow{\mathcal{L}} \mathcal{N}(0, I(\theta^*)).$$

In the next step we prove the convergence of  $\int_0^1 \frac{1}{t} \partial_{\bar{\theta}}^2 \tilde{\ell}_t(\theta^* + s(\bar{\theta}_t - \theta^*)) ds$ . From (5.9) we see that for  $(i, j) \in \{1, \dots, d\}$ ,

$$(5.12) \quad \begin{aligned} \partial_{\theta_i, \theta_j}^2 \tilde{\ell}_t(\theta) &= - \int_0^t \frac{(b(\theta, X_s) - b(\theta^*, X_s)) \partial_{\theta_i, \theta_j}^2 b(\theta, X_s)}{\sigma^2(X_s)} ds - \int_0^t \frac{\partial_{\theta_i} b(\theta, X_s) \partial_{\theta_j} b(\theta, X_s)}{\sigma^2(X_s)} ds \\ &\quad + \int_0^t \frac{\partial_{\theta_i, \theta_j}^2 b(\theta, X_s)}{\sigma(X_s)} dW_s \\ &:= U_t^1(\theta) + U_t^2(\theta) + U_t^3(\theta). \end{aligned}$$

Using the ergodic theorem,  $P$ -a.s.

$$\begin{aligned} \frac{1}{t} U_t^1(\theta) &\rightarrow U_{\infty}^1(\theta) := - \int_{\mathbb{R}} \frac{(b(\theta, x) - b(\theta^*, x)) \partial_{\theta_i, \theta_j}^2 b(\theta, x)}{\sigma^2(x)} \pi(dx); \\ \frac{1}{t} U_t^2(\theta) &\rightarrow U_{\infty}^2(\theta) := - \int_{\mathbb{R}} \frac{\partial_{\theta_i} b(\theta, x) \partial_{\theta_j} b(\theta, x)}{\sigma^2(x)} \pi(dx) = -I_{i,j}(\theta). \end{aligned}$$

Moreover, using Assumption 7 and 8 and the same argument which were used to prove the equicontinuity (5.5) we obtain that the families of functions  $(\theta \mapsto \frac{U_t^1}{t}(\theta))_{t \geq 0}$  and  $(\theta \mapsto \frac{U_t^2}{t}(\theta))_{t \geq 0}$  are almost surely equicontinuous. Finally, the uniform law of large numbers for local martingales [15] together with Assumptions 5, 7 and 8 gives that  $P$ -a.s.

$$\sup_{\theta \in \Theta} t^{-1} |U_t^3(\theta)| = \sup_{\theta \in \Theta} t^{-1} \left| \int_0^t \frac{\partial_{\bar{\theta}}^2 b(\theta, X_s)}{\sigma(X_s)} dW_s \right| \rightarrow 0.$$

Using (5.12) and the three last displays we obtain  $P$ -a.s.

$$\sup_{\theta \in \Theta} \left| t^{-1} \partial_{\theta}^2 \tilde{\ell}_t(\theta) - (U_{\infty}^1(\theta) - I(\theta)) \right| \rightarrow 0.$$

Using this uniformity together with the a.s. convergence  $\bar{\theta}_t \rightarrow \theta^*$  we get  $P$ -a.s.

$$\sup_{s \in [0,1]} \left| t^{-1} \partial_{\theta}^2 \tilde{\ell}_t(\theta^* + s(\bar{\theta}_t - \theta^*)) - (-I(\theta^*)) \right| \rightarrow 0$$

and

$$(5.13) \quad t^{-1} \int_0^1 \partial_{\theta}^2 \tilde{\ell}_t(\theta^* + s(\bar{\theta}_t - \theta^*)) ds \rightarrow -I(\theta^*).$$

Finally, from the non-degeneracy of the Fisher information matrix  $I(\theta^*)$ , (5.10), (5.11), (5.13), and Slutsky's theorem, we deduce the asymptotic normality of the estimator.  $\square$

**5.2. Local asymptotic normality and efficiency.** To obtain an asymptotic efficiency result in the sense of Hájek-Le Cam's convolution theorem we state now the local asymptotic normality property for the statistical experiment induced by the family  $(P^{\theta})_{\theta \in \Theta}$ . From this result we can deduce the efficiency of the discretized estimator with jump filter (cf. Theorem 3.2).

**THEOREM 5.3.** *Suppose that Assumptions 1 to 9 are satisfied. Then the family  $(P^{\theta})_{\theta \in \Theta}$  is locally asymptotically normal. That is, for all  $h \in \mathbb{R}^d$ , we have the convergence in distribution under  $P$ ,*

$$(5.14) \quad \ell_t(\theta^* + \frac{h}{\sqrt{t}}) - \ell_t(\theta^*) \xrightarrow{\mathcal{L}} -1/2 h^{\top} I(\theta^*) h + N, \quad \text{as } t \rightarrow \infty,$$

where  $N \sim \mathcal{N}(0, h^{\top} I(\theta^*) h)$ . As a consequence the drift estimator  $\bar{\theta}_t$  is asymptotically efficient in the sense of the Hájek-Le Cam convolution theorem.

The proof of Theorem 5.3 is given in Section 8 (see [Supplement](#)).

**5.3. Proofs of Theorems 3.1 and 3.2.**

**PROOF OF THEOREM 3.1.** Let  $\tilde{\ell} : \Theta \rightarrow \mathbb{R}$  be given by (5.4) and define

$$(5.15) \quad \ell(\theta) = \tilde{\ell}(\theta) + \frac{1}{2} \pi \left( \frac{b^2(\theta^*, x)}{\sigma^2(x)} \right).$$

Under Assumptions 1 and 5 and Lemma 2.1 the last term in the right hand side of (5.15) is finite. We will apply Wald's method for proving consistency of  $M$  estimators (see for example

Theorem 5.7 in [30]). It follows from (5.8) that  $\sup_{\theta; d(\theta, \theta^*) \geq \varepsilon} \ell(\theta) < \ell(\theta^*)$ . Therefore, it remains to prove that  $\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |t_n^{-1} \ell_{t_n}^n(\theta) - \ell(\theta)| = 0$  in probability. To obtain this last statement we decompose this difference as follows:

$$(5.16) \quad \sup_{\theta \in \Theta} |\ell(\theta) - t_n^{-1} \ell_{t_n}^n(\theta)| \leq \sup_{\theta \in \Theta} |\ell(\theta) - t_n^{-1} \ell_{t_n}(\theta)| + \sup_{\theta \in \Theta} |t_n^{-1} (\ell_{t_n}(\theta) - \ell_{t_n}^n(\theta))|.$$

Using respectively the Ergodic Theorem given by Lemma 2.1 (1) and the Law of Large Numbers for continuous local martingales ([24] p.178) we see that a.s.

$$\frac{1}{t} \int_0^t \frac{b^2(\theta^*, X_s)}{\sigma^2(X_s)} ds \rightarrow \pi \left( \frac{b^2(\theta^*, x)}{\sigma^2(x)} \right) \quad \text{and} \quad \frac{1}{t} \int_0^t \frac{b(\theta^*, X_s)}{\sigma(X_s)} dW_s \rightarrow 0.$$

Using these last display, (5.3), and (5.7), we see that the first term of the decomposition (5.16) tends to zero  $P$ -almost surely. In order to show the convergence to zero in probability of the second term, we decompose it as follows, using (3.4),

$$\sup_{\theta \in \Theta} |t_n^{-1} (\ell_{t_n}(\theta) - \ell_{t_n}^n(\theta))| \leq \sup_{\theta \in \Theta} t_n^{-1} |J_n^1(\theta)| + \sup_{\theta \in \Theta} t_n^{-1} |J_n^2(\theta)|,$$

where we have denoted

$$J_n^1(\theta) := \int_0^{t_n} \sigma(X_s)^{-2} b(\theta, X_s) dX_s^c - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} b(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i}$$

and

$$J_n^2(\theta) := \frac{1}{2} \int_0^{t_n} \sigma(X_s)^{-2} b(\theta, X_s)^2 ds - \frac{1}{2} \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} b(\theta, X_{t_{i-1}})^2 \Delta_i^n Id.$$

Hence, it remains to prove the convergence to zero of  $t_n^{-1} |J_n^1(\theta)|$  and  $t_n^{-1} |J_n^2(\theta)|$  uniformly in  $\theta$ . For  $t_n^{-1} |J_n^1(\theta)|$  we apply Proposition 4.1, together with the fact that  $n\Delta_n = O(t_n)$ . Indeed, using Assumptions 7 and 8 we see that the function  $f(\theta, x) = \sigma(x)^{-2} b(\theta, x)^2$  satisfies all the assumptions of the Proposition 4.1. For the second term  $t_n^{-1} |J_n^2(\theta)|$  we use Lemma 4.4.  $\square$

PROOF OF THEOREM 3.2. A Taylor expansion around  $\hat{\theta}_n$  yields

$$(5.17) \quad \frac{1}{t_n} \int_0^1 \partial_{\theta}^2 \ell_{t_n}^n(\theta^* + s(\hat{\theta}_n - \theta^*)) ds \times t_n^{1/2} (\hat{\theta}_n - \theta^*) = -\frac{1}{t_n^{1/2}} \nabla_{\theta} \ell_{t_n}^n(\theta^*).$$

For the right hand side we write that

$$(5.18) \quad \frac{1}{t_n^{1/2}} \nabla_{\theta} \ell_{t_n}^n(\theta^*) = \frac{\nabla_{\theta} \ell_{t_n}^n(\theta^*) - \nabla_{\theta} \ell_{t_n}(\theta^*)}{t_n^{1/2}} + \frac{\nabla_{\theta} \ell_{t_n}(\theta^*)}{t_n^{1/2}}.$$



By (5.11) we have that under  $P$

$$(5.19) \quad \frac{\nabla_{\theta} \ell_{t_n}(\theta^*)}{t_n^{1/2}} \xrightarrow{\mathcal{L}} N(0, I(\theta^*)), \quad n \rightarrow \infty.$$

The first term of the sum on the right hand side of (5.18) has the form

$$\begin{aligned} & \frac{\nabla_{\theta} \ell_{t_n}^n(\theta^*) - \nabla_{\theta} \ell_{t_n}(\theta^*)}{t_n^{1/2}} \\ &= -t_n^{-1/2} \left( \int_0^{t_n} \sigma(X_s)^{-2} \nabla_{\theta} b(\theta^*, X_s) dX_s^c - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \nabla_{\theta} b(\theta^*, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right) \\ &+ t_n^{-1/2} \frac{1}{2} \left( \int_0^{t_n} \sigma(X_s)^{-2} \nabla_{\theta} b(\theta^*, X_s)^2 ds - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \nabla_{\theta} b(\theta^*, X_{t_{i-1}})^2 \Delta_i^n Id \right). \end{aligned}$$

By applying Proposition 4.1 for  $k = 1, \dots, d$  with  $f_k(\theta^*, x) = \sigma(x)^{-2} \partial_{\theta_k} b(\theta^*, x)$ , and using Assumptions 7–8 we obtain that

$$t_n^{-1/2} \left( \int_0^{t_n} \sigma(X_s)^{-2} \partial_{\theta_k} b(\theta^*, X_s) dX_s^c - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \partial_{\theta_k} b(\theta^*, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right) \xrightarrow{P} 0$$

as  $n \rightarrow \infty$ . Furthermore, Lemma 4.4 (ii) leads to

$$t_n^{-1/2} \left( \int_0^{t_n} \sigma(X_s)^{-2} \partial_{\theta_k} b(\theta^*, X_s)^2 ds - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \partial_{\theta_k} b(\theta^*, X_{t_{i-1}})^2 \Delta_i^n Id \right) \xrightarrow{P} 0,$$

as  $n \rightarrow \infty$ . Combining now the last three displays results in  $\frac{\nabla_{\theta} \ell_{t_n}^n(\theta^*) - \nabla_{\theta} \ell_{t_n}(\theta^*)}{t_n^{1/2}} \xrightarrow{P} 0$  such

that (5.18) and (5.19) give  $t_n^{-1/2} \nabla_{\theta} \ell_{t_n}^n(\theta^*) \xrightarrow{d} N(0, I(\theta^*))$ ,  $n \rightarrow \infty$ .

To finish the proof it remains to show the convergence of the left hand side in (5.17). For  $(j, k) \in \{1, \dots, d\}^2$  and  $\theta \in \Theta$ ,

$$\begin{aligned} & t_n^{-1} \sup_{\theta \in \Theta} \left| \left( \partial_{\theta_j \theta_k}^2 \ell_{t_n}^n(\theta) - \partial_{\theta_j \theta_k}^2 \ell_{t_n}(\theta) \right) \right| \\ & \leq t_n^{-1} \sup_{\theta \in \Theta} \left| \int_0^{t_n} \sigma(X_s)^{-2} \partial_{\theta_j \theta_k}^2 b(\theta, X_s) dX_s^c - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \partial_{\theta_j \theta_k}^2 b(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right| \\ & + t_n^{-1} \sup_{\theta \in \Theta} \left| \int_0^{t_n} \sigma(X_s)^{-2} \partial_{\theta_j} b(\theta, X_s) \partial_{\theta_k} b(\theta, X_s) ds - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \partial_{\theta_j} b(\theta, X_s) \partial_{\theta_k} b(\theta, X_s) \Delta_i^n Id \right| \\ & + t_n^{-1} \sup_{\theta \in \Theta} \left| \int_0^{t_n} \sigma(X_s)^{-2} \partial_{\theta_j \theta_k}^2 b(\theta, X_s) b(\theta, X_s) ds - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \partial_{\theta_j \theta_k}^2 b(\theta, X_s) b(\theta, X_s) \Delta_i^n Id \right| \\ & := U_n^1 + U_n^2 + U_n^3. \end{aligned}$$

Proposition 4.1 together with Assumptions 7–8 state that

$$(5.20) \quad U_n^1 \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty.$$

Lemma 4.4 (i) gives for  $k = 2, 3$

$$(5.21) \quad U_n^k \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty.$$

Recall that, by (5.1),  $\bar{\theta}_t$  is a maximizer of the log-likelihood (3.2) based on the continuous observations. Combining (5.20) and (5.21) with consistency of  $\hat{\theta}$  and  $\bar{\theta}$  we get

$$\int_0^1 \frac{1}{t_n} |\partial_{\bar{\theta}}^2 \ell_{t_n}^n(\theta^* + s(\hat{\theta}_n - \theta^*)) - \partial_{\bar{\theta}}^2 \ell_{t_n}(\theta^* + s(\bar{\theta}_{t_n} - \theta^*))| ds \xrightarrow{P} 0,$$

and hence, using (5.13), we deduce  $\frac{1}{t_n} \int_0^1 \partial_{\bar{\theta}}^2 \ell_{t_n}^n(\theta^* + s(\hat{\theta}_n - \theta^*)) ds \xrightarrow{P} -I(\theta^*)$ , as  $n \rightarrow \infty$ . The result follows.  $\square$

**6. Proofs for jump filtering.** In this section we prove the results stated in Section 4. We start by proving Lemma 4.2. We recall some notations:  $\mu$  denotes the Poisson random measure on  $[0, \infty) \times \mathbb{R}$  associated with the jumps of the Lévy process  $L$ , the intensity of this jump measure is  $ds \times \nu(dz)$ . We define  $\tilde{\mu} = \mu - ds \times \nu(dz)$  as the compensated Poisson measure, and we have  $L_t = \int_0^t \int_{\mathbb{R}} z \mu(ds, dz)$ .

In the proof we use some moment inequalities for jump diffusions and their continuous parts, gathered in the following lemma.

LEMMA 6.1. *Let  $X$  satisfy Assumptions 1–4. Let  $\mathcal{F}_s = \sigma\{(W_u)_{0 < u \leq s}, (L_u)_{0 < u \leq s}, X_0\}$ . Then for all  $t > s$ ,*

- (1) for all  $p \geq 2$ ,  $E[|X_t - X_s|^p]^{1/p} \leq C|t - s|^{1/p}$ ,
- (2) for all  $p \geq 2$ ,  $p \in \mathbb{N}$ ,  $E[|X_t - X_s|^p | \mathcal{F}_s] \leq C|t - s|(1 + |X_s|^p)$ ,
- (3) for all  $p > 1$ ,  $E[|X_t^c - X_s^c|^p]^{1/p} \leq C|t - s|^{1/2}$ .

The proof of Lemma 6.1 is given in Section 8 (see Supplement).

REMARK 2. *Using Lemma 6.1 (3), the definition of  $v_n^i$  (3.3) and Markov's inequality we see that for all  $p > 1$ ,*

$$P(|\Delta_i^n X^c| \geq v_n^i) \leq P(|\Delta_i^n X^c| \geq \underline{a}v_n) = O(\Delta_n^{p/2} v_n^{-p}) = O(\Delta_n^{\varepsilon p}).$$

In our proofs we extensively use the Lemma 9 from Genon-Catalot and Jacod [8], that we recall below.

LEMMA 6.2 ([8]). Let  $(\mathcal{F}_i^n), i = 1 \dots, n$ , be a filtration for each  $n \geq 1$ ,  $(e_i^n), i = 1, \dots, n; n \in \mathbb{N}$ ,  $U$  be random variables, with  $e_i^n$  being  $\mathcal{F}_i^n$  measurable. Suppose that

$$\sum_{i=1}^n E[e_i^n | \mathcal{F}_{i-1}^n] \xrightarrow{P} U \quad \text{and} \quad \sum_{i=1}^n E[(e_i^n)^2 | \mathcal{F}_{i-1}^n] \xrightarrow{P} 0.$$

Then,  $\sum_{i=1}^n e_i^n \rightarrow U$  in probability.

PROOF OF LEMMA 4.2. We start by proving (i). For all  $n \in \mathbb{N}^*$ ,  $i \in \mathbb{N}^*$  we define the set where increments of  $X$  are small:

$$K_n^i = \{|\Delta_i^n X| \leq v_n^i\},$$

and the event on which all the jumps of  $L$  on the interval  $(t_{i-1}, t_i]$  are small :

$$(6.1) \quad N_n^i = \{|\Delta L_s| \leq 3v_n^i/\gamma_{min}; \forall s \in (t_{i-1}, t_i]\}, \text{ where } \Delta L_s := L_s - L_{s-}.$$

Let us denote by  $X^J$  the jump part of  $X$  given by

$$X_t^J = \int_0^t \int_{\mathbb{R} \setminus \{0\}} \gamma(X_{s-}) z \mu(ds, dz), \quad t \geq 0.$$

Using these notations, we introduce the quantities

$$\begin{aligned} B_n^1(\theta) &:= - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^J) 1_{K_n^i \cap (N_n^i)^c}; & B_n^2(\theta) &:= - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^J) 1_{K_n^i \cap N_n^i}; \\ B_n^3(\theta) &:= \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^c) 1_{(K_n^i)^c \cap (N_n^i)^c}; & B_n^4(\theta) &:= \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^c) 1_{(K_n^i)^c \cap N_n^i}. \end{aligned}$$

We can decompose the difference as follows

$$(6.2) \quad \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left( \Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right) = B_n^1(\theta) + B_n^2(\theta) + B_n^3(\theta) + B_n^4(\theta).$$

We start by studying the convergence of  $B_n^1(\theta)$ . The proof for the control of this term is slightly different in the case where Assumption 4 (i)–(iv) holds and in the case where Assumption 4 (i),(ii') holds. We focus first on the former case, as it contains the situation  $\nu(\mathbb{R}) = \infty$  which is more difficult to address, and hence we assume now on that Assumption 4 (i)–(iv) holds.

Let  $T_i^* \in (t_{i-1}, t_i]$  such that  $|\Delta L_{T_i^*}| = \max\{|\Delta L_s|; s \in (t_{i-1}, t_i]\}$ . Remark that  $T_i^*$  is well defined, as from Assumption 4 (iii) there is, almost surely, a unique time at which the Lévy process admits a jump with maximal size. We introduce the event

$$(6.3) \quad A_n^i = \left\{ \sum_{t_{i-1} < s \leq t_i; s \neq T_i^*} |\Delta L_s| \leq \frac{v_n^i}{\gamma_{max}} \right\},$$

where  $\gamma_{max}$  is defined in Assumption 4 (iv).

To estimate  $B_n^1(\theta)$  we make the decomposition

$$K_n^i \cap (N_n^i)^c = K_n^i \cap (N_n^i)^c \cap A_n^i \quad \cup \quad K_n^i \cap (N_n^i)^c \cap (A_n^i)^c.$$

Note that

$$\begin{aligned} & K_n^i \cap (N_n^i)^c \cap A_n^i \\ & \subset \{|\Delta_i^n X^c + \gamma(X_{T_i^* -})\Delta L_{T_i^*}| \leq v_n^i; \quad |\gamma(X_{T_i^* -})\Delta L_{T_i^*}| > 3v_n^i; \quad \mid \sum_{\substack{t_{i-1} < s \leq t_i \\ s \neq T_i^*}} \Delta X_s \mid \leq v_n^i\} \\ & \subset \{|\Delta_i^n X^c| \geq v_n^i\} \subset \{|\Delta_i^n X^c| \geq \underline{a}v_n\}. \end{aligned}$$

Hence, using Remark 2 we get for all  $p > 1$ :

$$(6.4) \quad P(K_n^i \cap (N_n^i)^c \cap A_n^i) \leq P(|\Delta_i^n X^c| \geq \underline{a}v_n) = O(\Delta_n^{\varepsilon p}).$$

Then, using the  $L^2$ -isometry for stochastic integral with respect to the compensated Poisson measure and the Jensen's inequality, we get

$$\begin{aligned} E|\Delta_i^n X^J|^2 & \leq 2E \left| \int_{t_{i-1}}^{t_i} \int_{\mathbb{R} \setminus \{0\}} \gamma(X_{s-}) z \tilde{\mu}(ds, dz) \right|^2 + 2E \left[ \int_{t_{i-1}}^{t_i} \int_{\mathbb{R} \setminus \{0\}} \gamma(X_s) z ds \nu(dz) \right]^2 \\ & \leq 2 \int_{t_{i-1}}^{t_i} \int_{\mathbb{R} \setminus \{0\}} E[\gamma^2(X_s)] z^2 ds \nu(dz) + 2 \int_{t_{i-1}}^{t_i} \int_{\mathbb{R} \setminus \{0\}} E[\gamma^2(X_s)] |z| ds \nu(dz) \int_{t_{i-1}}^{t_i} \int_{\mathbb{R} \setminus \{0\}} |z| ds \nu(dz) \\ & = O(\Delta_n), \end{aligned}$$

where in the last line we have used Assumption 1, Assumption 3 (i), Assumption 4 (ii) and Lemma 2.1 statement (3).

Together with this last bound, Hölder's inequality, sub-polynomial growth of  $f$  and (3) from Lemma 2.1 this gives for any  $p > 1$  that

$$(6.5) \quad E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^J) \left( 1_{K_n^i \cap (N_n^i)^c \cap A_n^i} \right) \right| = O(n \Delta_n^{\varepsilon p}).$$

To estimate the contribution of the event  $K_n^i \cap (N_n^i)^c \cap (A_n^i)^c$  we need the following Lemma:

LEMMA 6.3. *Under assumptions 1–3 and 4(i)–(iv), we have for some  $C > 0$ ,*

$$P((N_n^i)^c \cap (A_n^i)^c) \leq C \frac{\Delta_n^2}{\underline{a}v_n / \gamma_{min}} \int_{|z| \geq 3\underline{a}v_n / \gamma_{min}} \nu(dz).$$

The proof of Lemma 6.3 is given in Section 8 (see [Supplement](#)).

Using Hölder inequality, sub-polynomial growth of  $f$ , Lemma 2.1 (3), and Lemma 6.3, we get for  $1/p + 1/q = 1$  and some  $C > 0$ ,

$$\begin{aligned}
& E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left( \Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right) \mathbf{1}_{K_n^i \cap (N_n^i)^c \cap (A_n^i)^c} \right| \\
& \leq \sum_{i=1}^n \left( E \sup_{\theta \in \Theta} |f(\theta, X_{t_{i-1}})|^p (|\Delta_i^n X^c| + v_n^i)^p \right)^{1/p} \left( P((N_n^i)^c \cap (A_n^i)^c) \right)^{1/q} \\
& \leq Cn(\Delta_n^{1/2} + \bar{a}v_n) \left( \frac{\Delta_n^2}{\underline{a}v_n} \int_{|z| \geq 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1/q}, \text{ using Lemma 6.1 (3),} \\
(6.6) \quad & \leq Cnv_n^{\varepsilon/2} \Delta_n^{2-\varepsilon} \left( \int_{|z| \geq 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2}, \text{ choosing } 1/q = 1 - \varepsilon/2.
\end{aligned}$$

From (6.5) and (6.6), we get

$$(6.7) \quad E \sup_{\theta \in \Theta} |B_n^1(\theta)| = o \left( n\Delta_n^{3/2-\varepsilon/2} \left( \int_{|z| \geq \underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2} \right) + O(n\Delta_n^2)$$

This gives a control for  $B_n^1(\theta)$  in the situation where Assumption 4 (i)–(iv) holds true. Actually (6.7) is valid if we replace Assumption 4 (i)–(iv) with Assumption 4 (i), (ii') using Lemma 6.4 below.

LEMMA 6.4. *Assume Assumptions 1–3, 4 (i), (ii'). Then, we have*

$$P(K_n^i \cap (N_n^i)^c) = O(\Delta_n^2) \text{ and, } E[\sup_{\theta \in \Theta} |B_n^1(\theta)|] = O(n\Delta_n^2).$$

The proof of Lemma 6.4 is given in Section 8 (see [Supplement](#)).

We now estimate  $B_n^2(\theta)$ . It is clear that  $E \sup_{\theta \in \Theta} |B_n^2(\theta)|$  is bounded by

$$\begin{aligned}
& \sum_{i=1}^n E \sup_{\theta \in \Theta} |f(\theta, X_{t_{i-1}})| \int_{t_{i-1}}^{t_i} \int_{\mathbb{R} \setminus \{0\}} \gamma(X_{s-}) z \mu(ds, dz) \mathbf{1}_{K_n^i \cap N_n^i} \\
& \leq \sum_{i=1}^n E \int_{t_{i-1}}^{t_i} \int_{|z| \leq 3v_n^i/\gamma_{\min}} \sup_{\theta \in \Theta} |f(\theta, X_{t_{i-1}})| \gamma(X_{s-}) |z| \mu(ds, dz) \\
& \leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \int_{|z| \leq 3\bar{a}v_n/\gamma_{\min}} E[\sup_{\theta \in \Theta} |f(\theta, X_{t_{i-1}})| \gamma(X_s)] |z| \nu(dz) ds = O(n\Delta_n) \int_{|z| \leq 3\bar{a}v_n/\gamma_{\min}} |z| \nu(dz).
\end{aligned}$$

Since  $\gamma_{min} \geq 1$ , we obtain,

$$(6.8) \quad E \sup_{\theta \in \Theta} |B_n^2(\theta)| = O(n\Delta_n \int_{|z| \leq 3\bar{a}v_n/\gamma_{min}} |z|\nu(dz)) \leq O(n\Delta_n \int_{|z| \leq 3\bar{a}v_n} |z|\nu(dz)).$$

To estimate  $B_n^3(\theta)$ , we first give a useful control on the probability of  $(N_n^i)^c$  :

$$(6.9) \quad \begin{aligned} P((N_n^i)^c) &= 1 - P\left(\int_{t_{i-1}}^{t_i} \int_{|z| > 3v_n^i/\gamma_{min}} \mu(ds, dz) = 0\right) \leq 1 - P\left(\int_{t_{i-1}}^{t_i} \int_{|z| > 3\underline{a}v_n/\gamma_{min}} \mu(ds, dz) = 0\right) \\ &\leq 1 - e^{-\Delta_n \int_{|z| > 3\underline{a}v_n/\gamma_{min}} \nu(dz)} = O\left(\Delta_n \int_{|z| > 3\underline{a}v_n/\gamma_{min}} \nu(dz)\right). \end{aligned}$$

Hence, using Hölder's inequality, the assumptions on  $f$  and (3) of Lemma 6.1 we obtain for any  $q > 1$  that

$$(6.10) \quad \begin{aligned} E \sup_{\theta \in \Theta} |B_n^3(\theta)| &\leq E \sum_{i=1}^n \sup_{\theta \in \Theta} |f(\theta, X_{t_{i-1}})| |\Delta_i^n X^c| 1_{\{|\Delta_i^n X| > v_n, (N_n^i)^c\}} \\ &\leq O(\Delta_n^{1/2}) \sum_{i=1}^n P((N_n^i)^c)^{1/q} \leq O(n\Delta_n^{1/2}) \Delta_n^{1/q} \left(\int_{|z| > 3\underline{a}v_n/\gamma_{min}} \nu(dz)\right)^{1/q}. \end{aligned}$$

It remains to estimate the term  $B_n^4(\theta)$  in the decomposition (6.2). Observe that for all  $p' > 1$ ,

$$\begin{aligned} P((K_n^i)^c \cap N_n^i) &= \\ P(|\Delta_i^n X^c + \sum_{t_{i-1} < s \leq t_i} \Delta X_s| > v_n^i; N_n^i) &\leq P(|\Delta_i^n X^c| > \frac{v_n^i}{2}) + P(|\sum_{t_{i-1} < s \leq t_i} \Delta X_s| > \frac{v_n^i}{2}; N_n^i) \leq \\ C\Delta_n^{\varepsilon p'} + P(|\int_{t_{i-1}}^{t_i} \gamma(X_{s-}) \int_{|z| \leq 3v_n^i/\gamma_{min}} z\mu(ds, dz)| > \frac{\underline{a}v_n}{2}) &\leq C\Delta_n^{\varepsilon p'} + \frac{C\Delta_n}{\underline{a}v_n} \int_{|z| \leq 3\bar{a}v_n/\gamma_{min}} |z|\nu(dz), \end{aligned}$$

where  $C > 0$ . Using Hölder's inequality twice, this last bound, sub-polynomial growth of  $f$

and Lemma 6.1 (3) we can easily see that with  $1/q = 1 - \varepsilon/2$  we get

$$\begin{aligned}
E \sup_{\theta \in \Theta} |B_n^4(\theta)| &= E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^c) 1_{(K_n^i)^c \cap N_n^i} \right| \\
&\leq \sum_{i=1}^n \left( E \sup_{\theta \in \Theta} |f(\theta, X_{t_{i-1}})|^p |\Delta_i^n X^c|^p \right)^{1/p} P^{1/q}((K_n^i)^c \cap N_n^i) \\
&\leq Cn \Delta_n^{1/2} \left( \Delta_n^{\varepsilon p'} + \frac{\Delta_n}{v_n} \int_{|z| \leq 3\bar{a}v_n/\gamma_{min}} |z| \nu(dz) \right)^{1-\varepsilon/2} \\
&\leq Cn \Delta_n^{1/2} \Delta_n^{(1/2+\varepsilon)(1-\varepsilon/2)} \left( \int_{|z| \leq 3\bar{a}v_n/\gamma_{min}} |z| \nu(dz) \right)^{1-\varepsilon/2} + Cn \Delta_n^2, \\
(6.11) \quad &\leq Cn \Delta_n^{1+\varepsilon/4} \left( \int_{|z| \leq 3\bar{a}v_n/\gamma_{min}} |z| \nu(dz) \right)^{1-\varepsilon/2} + Cn \Delta_n^2,
\end{aligned}$$

since  $p'$  can be chosen arbitrarily large. Finally, collecting (6.7), (6.8), (6.10) with  $1/q = 1 - \varepsilon/2$ , and (6.11), and using  $\gamma_{min} \leq 1$ , we obtain first part of the lemma.

We continue with the proof of (ii). In the (i) we obtain a bound which is used in the proof of the consistency. This bound vanishes when divided by  $n\Delta_n$ . The bound of (ii) is used to prove the asymptotic normality of the estimator, with rate  $\sqrt{n\Delta_n}$ , and is sharper. Especially, the control (6.10) we have obtained for the term  $\sup_{\theta \in \Theta} |B_n^3(\theta)|$  is insufficient. Below, we find an improved control on  $|B_n^3(\theta)|$ , but without the uniformity with respect to  $\theta$ .

To prove (ii), we consider again the decomposition (6.2). Using (6.5) and (6.6) we can see that

$$E|B_n^1(\theta)| = o \left( n \Delta_n^{2-\varepsilon} \left( \int_{|z| \geq 3\bar{a}v_n/\gamma_{min}} \nu(dz) \right)^{1-\varepsilon/2} \right) + O(n \Delta_n^2),$$

while (6.8) gives the bound for  $E|B_n^2(\theta)|$ , and (6.11) gives the bound for  $E|B_n^4(\theta)|$ . Hence, proving the point (ii) of the lemma reduces to show that

$$B_n^3(\theta) = o_P(\sqrt{n\Delta_n}) + o_{L^1} \left( n \Delta_n^{2-\varepsilon} \left( \int_{|z| \geq 3\bar{a}v_n/\gamma_{min}} \nu(dz) \right)^{1-\varepsilon/2} \right) + O_{L^1}(n \Delta_n^2).$$

To estimate  $B_n^3(\theta)$  we use a decomposition

$$(6.12) \quad B_n^3(\theta) = \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^c) 1_{(N_n^i)^c} - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^c) 1_{K_n^i \cap (N_n^i)^c}.$$

We will show that the first term of this decomposition goes to zero after suitable normalization. Let  $e_i := f(\theta, X_{t_{i-1}})\Delta_n^i X^c \mathbf{1}_{(N_n^i)^c}$ . Denote  $\mathcal{F}_i^n = \mathcal{F}_{t_i} = \sigma\{(W_s)_{0 \leq s \leq t_i}, (L_s)_{0 \leq s \leq t_i}, X_0\}$ , then

$$\begin{aligned} E[e_i | \mathcal{F}_{i-1}^n] &= f(\theta, X_{t_{i-1}}) E \left[ \int_{t_{i-1}}^{t_i} \sigma(X_s) dW_s \mathbf{1}_{(N_n^i)^c} | \mathcal{F}_{i-1}^n \right] \\ &\quad + f(\theta, X_{t_{i-1}}) E \left[ \int_{t_{i-1}}^{t_i} b(\theta^*, X_s) ds \mathbf{1}_{(N_n^i)^c} | \mathcal{F}_{i-1}^n \right]. \end{aligned}$$

Observe that  $(W_s)_{s \geq 0}$  remains a Brownian motion with respect to the filtration enlarged by  $\sigma((L_s)_{s \geq 0})$ , since  $L$  and  $W$  are independent. Therefore,

$$E \left[ \int_{t_{i-1}}^{t_i} \sigma(X_s) dW_s \mathbf{1}_{(N_n^i)^c} | \mathcal{F}_{i-1}^n \right] = E \left[ \mathbf{1}_{(N_n^i)^c} E \left[ \int_{t_{i-1}}^{t_i} \sigma(X_s) dW_s | \mathcal{F}_{i-1}^n \vee \sigma((L_s)_{s \geq 0}) \right] | \mathcal{F}_{i-1}^n \right] = 0,$$

where we have used that  $N_n^i$  is  $\mathcal{F}_{i-1}^n \vee \sigma((L_s)_{s \geq 0})$  measurable, recalling (6.1) and the fact that  $v_n^i$  is  $\mathcal{F}_{i-1}^n$  measurable. This yields to

$$|E[e_i | \mathcal{F}_{i-1}^n]| \leq |f(\theta, X_{t_{i-1}})| \int_{t_{i-1}}^{t_i} E [|b(\theta^*, X_s)| \mathbf{1}_{(N_n^i)^c} | \mathcal{F}_{i-1}^n] ds.$$

With a proof similar to the proof of (6.9), we can show that

$$(6.13) \quad P((N_n^i)^c | \mathcal{F}_{i-1}^n) = O \left( \Delta_n \int_{|z| > 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right).$$

Then, using Hölder inequality, Lipschitz continuity of  $b(\theta^*, \cdot)$  we write for  $p, q$  such that  $p^{-1} + q^{-1} = 1$ ,  $p \geq 2$  and  $C > 0$ ,

$$\begin{aligned} E [|b(\theta^*, X_s)| \mathbf{1}_{(N_n^i)^c} | \mathcal{F}_{i-1}^n] &\leq (E [|b(\theta^*, X_s)|^p | \mathcal{F}_{i-1}^n])^{1/p} P((N_n^i)^c | \mathcal{F}_{i-1}^n)^{1/q} \\ &\leq C (E [|b(\theta^*, X_s) - b(\theta^*, X_{t_{i-1}})|^p | \mathcal{F}_{t_{i-1}}] + |b(\theta^*, X_{t_{i-1}})|^p)^{1/p} P((N_n^i)^c | \mathcal{F}_{i-1}^n)^{1/q} \\ &\leq C \left( (E [|X_s - X_{t_{i-1}}|^p | \mathcal{F}_{t_{i-1}}])^{1/p} + |b(\theta^*, X_{t_{i-1}})| \right) (\Delta_n \int_{|z| > 3\underline{a}v_n/\gamma_{\min}} \nu(dz))^{1/q} \\ &\leq C (\Delta_n \int_{|z| > 3\underline{a}v_n/\gamma_{\min}} \nu(dz))^{1/q} \left( \Delta_n^{1/p} (1 + |X_{t_{i-1}}|^p)^{1/p} + |b(\theta^*, X_{t_{i-1}})| \right), \end{aligned}$$

where in the last line we have used Lemma 6.1 (2).

Using the fact that  $b(\theta^*, \cdot)$  and  $|f(\theta, \cdot)|$  are sub-polynomial and choosing again  $1/q = 1 - \varepsilon/2$ , (which also guarantees  $p > 2$ ), we obtain

$$|E[e_i | \mathcal{F}_{i-1}^n]| \leq h(|X_{t_{i-1}}|) \Delta_n^{2-\varepsilon/2} \left( \int_{|z| > 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2},$$



where  $h$  is a polynomial. Hence, under the condition  $n\Delta_n^{3-\varepsilon} \left( \int_{|z|>3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{2-\varepsilon} \rightarrow 0$ ,

$$(6.14) \quad E \left[ \sum_{i=1}^n \left| E \left[ \frac{e_i}{\sqrt{n\Delta_n}} \middle| \mathcal{F}_{i-1}^n \right] \right| \right] = O \left( n^{1/2} \Delta_n^{3/2-\varepsilon/2} \left( \int_{|z|>3\underline{a}\frac{v_n}{\gamma_{\min}}} \nu(dz) \right)^{1-\varepsilon/2} \right) \rightarrow 0.$$

Next, we bound the second moment of  $e_i$ . By Hölder's inequality with  $1/q = 1 - \varepsilon/2$ ,  $1/p = 1 - 1/q$ , , the sub-linear growth of  $f$ , together with Lemma 6.1 (3), we have

$$\begin{aligned} E[e_i^2] &\leq E[f(\theta, X_{t_{i-1}})^{2p} (\Delta_i^n X^c)^{2p}]^{1/p} P[(N_n^i)^c]^{1/q} \\ &\leq C \Delta_n P[(N_n^i)^c]^{1-\varepsilon/2} = O \left( \Delta_n^{2-\varepsilon/2} \left( \int_{|z|>3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2} \right), \text{ by (6.13)}. \end{aligned}$$

Hence, using  $\Delta_n \int_{|z|>3\underline{a}v_n/\gamma_{\min}} \nu(dz) \leq C \frac{\Delta_n}{v_n} \int_{|z|>3\underline{a}v_n/\gamma_{\min}} |z| \nu(dz) \rightarrow 0$ , since  $\int_{|z|>0} |z| \nu(dz) < \infty$ , we have

$$(6.15) \quad \begin{aligned} E \left[ \left| \sum_{i=1}^n E \left[ \left( \frac{e_i}{\sqrt{n\Delta_n}} \right)^2 \middle| \mathcal{F}_i^n \right] \right| \right] &= \sum_{i=1}^n E \left[ \left( \frac{e_i}{\sqrt{n\Delta_n}} \right)^2 \right] \\ &= O \left( \Delta_n^{1-\varepsilon/2} \left( \int_{|z|>3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2} \right) \rightarrow 0. \end{aligned}$$

Under (6.14) and (6.15) we obtain from Lemma 9 in [8] (Lemma 6.2) that

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X^c \mathbf{1}_{(N_n^i)^c} = \sum_{i=1}^n \frac{e_i}{\sqrt{n\Delta_n}} \xrightarrow{P} 0$$

if  $n\Delta_n^{3-\varepsilon} \left( \int_{|z|>3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{2-\varepsilon} \rightarrow 0$ .

Recall that the second term in the decomposition (6.12) of  $B_n^3(\theta)$  is given by

$$\sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X^c \mathbf{1}_{K_n^i \cap (N_n^i)^c}.$$

We will now bound this term in  $\mathbb{L}^1$ . We first assume that Assumption 4 (i)–(iv) holds. Using the set  $A_n^i$  defined by (6.3) we decompose

$$\mathbf{1}_{K_n^i \cap (N_n^i)^c} = \mathbf{1}_{K_n^i \cap (N_n^i)^c \cap A_n^i} + \mathbf{1}_{K_n^i \cap (N_n^i)^c \cap (A_n^i)^c}.$$

Using (6.4) we deduce for all  $p > 1$ ,

$$E \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^c) \left( \mathbf{1}_{K_n^i \cap (N_n^i)^c \cap A_n^i} \right) \right| = O(n\Delta_n^{\varepsilon p}).$$

Then, exactly as for the proof of (6.6), we get

$$E \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^c) \left( 1_{K_n^i \cap (N_n^i)^c \cap (A_n^i)^c} \right) \right| = o \left( n \Delta_n^{2-\varepsilon} \left( \int_{|z| \geq 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2} \right).$$

As a result,

$$B_n^3(\theta) = o_P(\sqrt{n\Delta_n}) + o_{\mathbb{L}^1} \left( n \Delta_n^{2-\varepsilon} \left( \int_{|z| \geq 3\underline{a}v_n/\gamma_{\min}} \nu(dz) \right)^{1-\varepsilon/2} \right) + O_{\mathbb{L}^1}(n\Delta_n^2).$$

This proves the assertion (ii) of the Lemma in the case where Assumption 4 (i)–(iv) holds.

If we replace Assumption 4 (i)–(iv), with Assumption 4 (i), (ii'), one can use Lemma 6.4 to get that  $B_n^3(\theta) = o_P(\sqrt{n\Delta_n}) + O_{\mathbb{L}^1}(n\Delta_n^{5/2-\varepsilon})$ , and the results follows.  $\square$

PROOF OF LEMMA 4.3. Using  $dX_s^c = b(\theta^*, X_s)ds + \sigma(X_s)dW_s$  we decompose the difference as

$$(6.16) \quad \int_0^{t_n} f(\theta, X_s) dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X^c = A_{n,1}(\theta) + A_{n,2}(\theta) + A_{n,3}(\theta),$$

where

$$(6.17) \quad \begin{aligned} A_{n,1}(\theta) &:= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (f(\theta, X_s) - f(\theta, X_{t_{i-1}})) \sigma(X_s) dW_s, \\ A_{n,2}(\theta) &:= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (f(\theta, X_s) - f(\theta, X_{t_{i-1}})) (b(\theta^*, X_s) - b(\theta^*, X_{t_{i-1}})) ds, \\ A_{n,3}(\theta) &:= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (f(\theta, X_s) - f(\theta, X_{t_{i-1}})) b(\theta^*, X_{t_{i-1}}) ds. \end{aligned}$$

Let us start by proving (ii). Let as previously  $\mathcal{F}_t = \sigma\{X_0, W_u, L_u; u \leq t\}, t \geq 0$ . Using martingale property and Itô's isometry of the stochastic integral together with the finite increments formula applied to  $f$ , we obtain

$$\begin{aligned} E[A_{n,1}^2(\theta)] &= E \left[ \sum_{i=1}^n \left( \int_{t_{i-1}}^{t_i} (f(\theta, X_s) - f(\theta, X_{t_{i-1}})) \sigma(X_s) dW_s \right)^2 \right] \\ &= E \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (f(\theta, X_s) - f(\theta, X_{t_{i-1}}))^2 \sigma^2(X_s) ds \\ &\leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} E [(X_s - X_{t_{i-1}})^2 f'^2(\theta, \tilde{x}) \sigma^2(X_s)] ds, \end{aligned}$$

where  $\tilde{x}$  is a point between  $X_s$  and  $X_{t_{i-1}}$ . Note that  $|\tilde{x}| \leq |X_s| + |X_{t_{i-1}}|$ . Using sub-polynomial growth of  $\sigma$  and  $\sup_{\theta} |f'(\theta, \cdot)|$ , Hölder's inequality, (3) of the Lemma 2.1 and (1) of the Lemma 6.1 yields

$$(6.18) \quad E [(X_s - X_{t_{i-1}})^2 f'^2(\theta, \tilde{x}) \sigma^2(X_s)] \leq CE [|X_s - X_{t_{i-1}}|^{2q}]^{1/q} \leq C \Delta_n^{1/q},$$

where  $q > 1$  and  $C$  is a positive constant. Hence, for all  $\theta \in \Theta$ ,  $E[A_{n,1}^2(\theta)] \leq Cn \Delta_n^{1+1/q}$ , and consequently, if  $1/q = 1 - \varepsilon/2$ ,

$$(6.19) \quad \frac{1}{\sqrt{n \Delta_n}} A_{n,1}(\theta) \xrightarrow{L^2} 0.$$

Using Lipschitz continuity of  $b$ , the assumptions on  $f$  and the same arguments than for obtaining (6.18), we get that  $E [|(f(\theta, X_s) - f(\theta, X_{t_{i-1}}))(b(\theta^*, X_s) - b(\theta^*, X_{t_{i-1}}))|] \leq C \Delta_n^{1/q}$  for  $s \in (t_{i-1}, t_i]$ . It follows immediately that

$$(6.20) \quad E[\sup_{\theta \in \Theta} |A_{n,2}(\theta)|] \leq Cn \Delta_n^{1+1/q}.$$

Hence, by choosing  $1/q = 1 - \varepsilon/2$  such that  $n \Delta_n^{1+2/q} = n \Delta_n^{3-\varepsilon} \rightarrow 0$  it follows that

$$(6.21) \quad \frac{1}{\sqrt{n \Delta_n}} \sup_{\theta \in \Theta} |A_{n,2}(\theta)| \xrightarrow{L^1} 0.$$

Observe that by Itô's formula  $A_{n,3}(\theta)$  can be written as

$$A_{n,3}(\theta) = a_n(\theta) + b_n(\theta) + c_n(\theta),$$

where

$$\begin{aligned} a_n(\theta) &= \sum_{i=1}^n b(\theta^*, X_{t_{i-1}}) \int_{t_{i-1}}^{t_i} ds \int_{t_{i-1}}^s f'(\theta, X_u) \sigma(X_u) dW_u, \\ b_n(\theta) &= \sum_{i=1}^n b(\theta^*, X_{t_{i-1}}) \int_{t_{i-1}}^{t_i} ds \int_{t_{i-1}}^s \left[ f'(\theta, X_u) b(\theta^*, X_u) + f''(\theta, X_u) \frac{1}{2} \sigma^2(X_u) \right] du, \\ c_n(\theta) &= \sum_{i=1}^n b(\theta^*, X_{t_{i-1}}) \int_{t_{i-1}}^{t_i} ds \sum_{\tau \in [t_{i-1}, s]} (f(\theta, X_{\tau}) - f(\theta, X_{\tau-})), \end{aligned}$$

where non-zero terms in the last sum corresponds to the jumps of  $X$ . Denote

$$\bar{e}_i^n := \frac{1}{\sqrt{n \Delta_n}} \int_{t_{i-1}}^{t_i} ds \int_{t_{i-1}}^s b(\theta^*, X_{t_{i-1}}) f'(\theta, X_u) \sigma(X_u) dW_u$$

Using martingale property of the stochastic integral with respect to  $W$  we obtain

$$E [\bar{e}_i^n | \mathcal{F}_{i-1}^n] = 0,$$

where  $\mathcal{F}_{i-1}^n = \mathcal{F}_{t_{i-1}}$ . Using Hölder's inequality and isometry property of the stochastic integral we get

$$\begin{aligned} E(E[(\bar{e}_i^n)^2 | \mathcal{F}_{i-1}^n]) &= E[(\bar{e}_i^n)^2] \leq \frac{1}{n} \int_{t_{i-1}}^{t_i} ds E \left( \int_{t_{i-1}}^s b(\theta^*, X_{t_{i-1}}) f'(\theta, X_u) \sigma(X_u) dW_u \right)^2 \\ &= \frac{1}{n} \int_{t_{i-1}}^{t_i} ds \int_{t_{i-1}}^s E[b^2(\theta^*, X_{t_{i-1}}) f'^2(\theta, X_u) \sigma^2(X_u)] du \leq C \frac{\Delta_n^2}{n}, \end{aligned}$$

where in the last inequality we have used the uniform in  $\theta$  sub-polynomial growth of  $f'$  and  $b$ , sub-linear growth of  $\sigma$  and Lemma 2.1(3). Therefore

$$E \sum_{i=1}^n E[(\bar{e}_i^n)^2 | \mathcal{F}_{i-1}^n] \leq C \Delta_n^2 \rightarrow 0 \quad \text{when } n \rightarrow \infty.$$

We conclude, using Lemma 9 in [8] (Lemma 6.2), that  $\forall \theta \in \Theta$ ,

$$(6.22) \quad \frac{1}{\sqrt{n\Delta_n}} a_n(\theta) = \sum_{i=1}^n \bar{e}_i^n \xrightarrow{P} 0.$$

Using again uniform in  $\theta$  sub-polynomial growth of  $b, f', f''$ , sub-linearity of  $\sigma$  and (3) of Lemma 2.1 we easily see that

$$(6.23) \quad E \sup_{\theta \in \Theta} |b_n(\theta)| \leq C n \Delta_n^2.$$

Let us now derive a bound for the jump term  $c_n$ .

$$\begin{aligned} (6.24) \quad E \sup_{\theta \in \Theta} |c_n(\theta)| &\leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} ds E \left[ \int_{t_{i-1}}^s \int_{\mathbb{R} \setminus \{0\}} |b(\theta^*, X_{t_{i-1}})| |f(\theta, X_{u-} + \gamma(X_{u-})z) - f(\theta, X_{u-})| \mu(du, dz) \right] \\ &\leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} ds \int_{t_{i-1}}^s du \int_{\mathbb{R} \setminus \{0\}} E |b(\theta^*, X_{t_{i-1}}) f'(\theta, \tilde{x}) \gamma(X_{u-})| |z| \nu(dz), \end{aligned}$$

where in the second inequality we used again the finite increments formula and denoted  $\tilde{x}$  the corresponding point between  $X_{u-}$  and  $X_u = X_{u-} + \gamma(X_{u-})z$ . Note that again  $|\tilde{x}| \leq |X_{u-}| + |X_u|$ . According to the Assumptions 3 (i), (iii) and the assumption b) of the lemma, the functions  $\gamma, b(\theta^*, \cdot)$  and  $\sup_{\theta} |f'(\theta, \cdot)|$  are sub-polynomial, and  $\nu(|z|) < \infty$ . Therefore, using (3) from Lemma 2.1 we have

$$\sup_{\theta \in \Theta} \int_{\mathbb{R} \setminus \{0\}} E |b(\theta^*, X_{t_{i-1}}) f'(\theta, \tilde{x}) \gamma(X_{u-})| |z| \nu(dz) < \infty.$$

This last inequality together with (6.24) gives

$$(6.25) \quad E \sup_{\theta \in \Theta} |c_n(\theta)| = O(n\Delta_n^2).$$

From (6.22), (6.23) and (6.25) we conclude that under condition  $n\Delta_n^{3-\varepsilon} \rightarrow 0$ ,

$$(6.26) \quad \frac{1}{\sqrt{n\Delta_n}} A_{n,3}(\theta) \xrightarrow{P} 0.$$

Finally, the previous display together with (6.19) and (6.21) proves (ii) of the lemma.

To prove the claim (i) we will again use the decomposition of the difference given by (6.16). Using the same arguments as in (6.18) and Lemma 6.1 (1), we get for some  $p > 1$ ,  $C > 0$  and  $\tilde{x}$  between  $X_s$  and  $X_{t_{i-1}}$ :

$$\begin{aligned} E \sup_{\theta \in \Theta} |A_{n,3}(\theta)| &\leq C \sum_{i=1}^n \int_{t_{i-1}}^{t_i} E [|f'(\theta, \tilde{x})(1 + |X_{t_{i-1}}|^p)| |X_s - X_{t_{i-1}}|] ds \leq \\ &C \sum_{i=1}^n \int_{t_{i-1}}^{t_i} E \left( |X_s - X_{t_{i-1}}|^2 \right)^{1/2} (E [|f'(\theta, \tilde{x})|^2 (1 + |X_{t_{i-1}}|^{2p})])^{1/2} ds \leq \\ &C \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \Delta_n^{1/2} ds \leq Cn\Delta_n^{3/2}. \end{aligned}$$

Hence

$$\frac{1}{n\Delta_n} \sup_{\theta \in \Theta} |A_{n,3}(\theta)| \xrightarrow{L^1} 0.$$

The bound (6.20) gives

$$\frac{1}{n\Delta_n} \sup_{\theta \in \Theta} |A_{n,2}(\theta)| \xrightarrow{L^1} 0.$$

From (6.19) we know that

$$\forall \theta \in \Theta, \quad \frac{1}{n\Delta_n} A_{n,1}(\theta) \xrightarrow{P} 0.$$

Let us prove that this convergence holds uniformly with respect to  $\theta$ . Denote  $\phi : [0, t_n] \rightarrow [0, t_n]$ ,  $\phi(s) = t_{i-1}$  if  $t_{i-1} \leq s < t_i$ ,  $i = 0, \dots, n-1$ , and define

$$M_n(\theta) := \frac{1}{t_n} A_{n,1}(\theta) = \frac{1}{t_n} \int_0^{t_n} (f(\theta, X_s) - f(\theta, X_{\phi(s)})) \sigma(X_s) dW_s.$$

Using Burkholder-Davis-Gundy inequality, Hölder continuity of  $f$ , sub-polynomial growth of its Hölder constant  $K$ , sub-linear growth of  $\sigma$  and the boundedness of moments of  $X$  given by (3) of Lemma 2.1 we find that for any  $p \geq 2$  and some  $C > 0$ ,

$$\begin{aligned}
E|M_n(\theta) - M_n(\theta')|^p &\leq |\theta - \theta'|^{\kappa p} \frac{C}{t_n^{p/2}} E \left( \frac{1}{t_n} \int_0^{t_n} (K^2(X_s) + K^2(X_{\phi(s)})) \sigma(X_s)^2 ds \right)^{p/2} \\
&\leq |\theta - \theta'|^{\kappa p} \frac{C}{t_n^{p/2+1}} \int_0^{t_n} E (K^2(X_s) + K^2(X_{\phi(s)}))^{p/2} \sigma(X_s)^p ds \leq C|\theta - \theta'|^{\kappa p}.
\end{aligned}$$

Choosing  $p > \frac{d}{\kappa}$  and using the Theorem 20 in the Appendix of [10] we obtain the convergence  $\frac{1}{n\Delta_n} \sup_{\theta \in \Theta} |A_{n,1}(\theta)| \xrightarrow{P} 0$ , and statement (i) follows.  $\square$

We have now collected all the tools to prove the convergence of the jump filter approximation towards integral functionals with respect to the  $X^c$  as stated in Proposition 4.1.

PROOF OF PROPOSITION 4.1. We decompose the difference as follows:

$$\begin{aligned}
(6.27) \quad &\left| \int_0^{t_n} f(\theta, X_s) dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right| \leq \\
&\left| \int_0^{t_n} f(\theta, X_s) dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X^c \right| + \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i} \right|
\end{aligned}$$

and prove first statement (i). Using Lemma 4.3 the first term of (6.27) divided by  $n\Delta_n$  goes to zero uniformly without any condition on  $\Delta_n$ .

Lemma 4.2 (i) and the fact that  $v_n = \Delta_n^{1/2-\varepsilon}$  gives

$$\begin{aligned}
&(n\Delta_n)^{-1} \sup_{\theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) (\Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n^i}) \right| = \\
&O_{\mathbb{L}^1} \left( \Delta_n^{1/2-\varepsilon/2} \left( \int_{|z| \geq \underline{a}v_n/\gamma_{min}} \nu(dz) \right)^{1-\varepsilon/2} + \left( \int_{|z| \leq 3\bar{a}v_n} |z| \nu(dz) \right)^{1-\varepsilon/2} + \Delta_n \right) = \\
&O_{\mathbb{L}^1} \left( \frac{\Delta_n^{1/2-\varepsilon/2}}{v_n^{1-\varepsilon/2}} \left( \int_{|z| \geq \underline{a}v_n/\gamma_{min}} |z| \nu(dz) \right)^{1-\varepsilon/2} + \left( \int_{|z| \leq 3\bar{a}v_n} |z| \nu(dz) \right)^{1-\varepsilon/2} + \Delta_n \right) \xrightarrow{P} 0,
\end{aligned}$$

where we have used that  $\int_{z \neq 0} |z| \nu(dz) < \infty$  by Assumptions 3–4. Hence statement (i) is proved.

Now we prove statement (ii). For any  $\theta \in \Theta$ , under the condition  $n\Delta_n^{3-\varepsilon} \rightarrow 0$ , the second statement of Lemma 4.3 gives the convergence to 0 of the first term in the decomposition (6.27), divided by  $\sqrt{n\Delta_n}$ . The convergence to 0 of the second term of (6.27), divided by  $\sqrt{n\Delta_n}$ , immediately follows from Lemma 4.2 and the conditions of (ii).  $\square$

PROOF OF LEMMA 4.4. Let us first prove (i). Using Lemma 6.1 (1), with some  $\tilde{x}$  between  $X_{t_{i-1}}$  and  $X_s$  in the third line below we obtain:

$$\begin{aligned} & E \sup_{\theta \in \Theta} \left| \int_0^{t_n} f(\theta, X_s) ds - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i Id \right| = E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} f(\theta, X_s) - f(\theta, X_{t_{i-1}}) ds \right| \\ & \leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} E \left[ \sup_{\theta \in \Theta} |f(\theta, X_s) - f(\theta, X_{t_{i-1}})| \right] ds \leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} E \left[ \sup_{\theta \in \Theta} |f'(\theta, \tilde{x})| |X_s - X_{t_{i-1}}| \right] ds \\ & \leq \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \left( E \sup_{\theta \in \Theta} |f'(\theta, \tilde{x})|^2 \right)^{1/2} (E |X_s - X_{t_{i-1}}|^2)^{1/2} ds \leq Cn \Delta_n^{3/2}. \end{aligned}$$

We now prove (ii). We find that

$$\int_0^{t_n} f(\theta, X_s) ds - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n Id = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (f(\theta, X_s) - f(\theta, X_{t_{i-1}})) ds,$$

and it is then apparent that this term can be treated exactly as the term  $A_{n,3}(\theta)$  given by the equation (6.17). Hence, from (6.26) (which requires the condition  $n\Delta_n^{3-\varepsilon} \rightarrow 0$ ) we have the result.  $\square$

## SUPPLEMENTARY MATERIAL

### Supplement: "Jump filtering and efficient drift estimation for Lévy-driven SDE's"

([link to the supplementary document](#); .pdf). The supplement contains the two additional Sections 7–8. In Section 7, we investigate the numerical performance of the estimator for finite sample. We consider the case of Ornstein–Uhlenbeck and 'Hyperbolic' diffusion models, with finite or infinite activity jump measure. We compare the results for different choices of the threshold constants  $a_i^n$  (recall (3.3)). In Section 8, we give a proof the Lemma 2.1 about the ergodicity of the process, and of the LAN property (Theorem 5.3). Next, we gather the proofs of the technical Lemmas 6.1, 6.3 and 6.4, and show that the identifiability Assumption 6 is equivalent to (2.1).

**Acknowledgements.** We are grateful to the anonymous referees for their very helpful comments and suggestions which improved the paper.

### References.

- [1] APPLEBAUM, D. (2009). *Lévy processes and stochastic calculus. 2nd ed.* Cambridge Studies in Advanced Mathematics 116. Cambridge University Press.
- [2] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **63** 167-241.
- [3] BIBINGER, M. and WINKELMANN, L. (2015). Econometrics of co-jumps in high-frequency data with noise. *Journal of Econometrics* **184** 361 - 378.

- [4] CONT, R. and TANKOV, P. (2004). *Financial modelling with jump processes*. Chapman & Hall.
- [5] DITLEVSEN, S. and GREENWOOD, P. (2013). The Morris–Lecar neuron model embeds a leaky integrate-and-fire model. *Journal of Mathematical Biology* **67** 239-259.
- [6] FIGUEROA-LÓPEZ, J. E. and NISEN, J. (2013). Optimality Thresholded realized power variations for Lévy jump diffusion models. *Stochastic Process. Appl.* **123** 2648-2677.
- [7] FLORENS-ZIMROU, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* **20** 547-557.
- [8] GENON-CATALOT, V. and JACOD, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques* **29** 119-151.
- [9] HUTTON, J. E. and NELSON, P. (1984). Interchanging the order of differentiation and stochastic integration. *Stochastic Processes and their Applications* **18** 371-377.
- [10] I. IBRAGIMOV, R. H. (2013). *Statistical estimation: asymptotic theory*.
- [11] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit theorems for stochastic processes. 2nd ed.* Grundlehren der Mathematischen Wissenschaften. 288. Berlin: Springer.
- [12] KESSLER, M. (1997). Estimation of an Ergodic Diffusion from Discrete Observations. *Scandinavian Journal of Statistics* **24** 211–229.
- [13] KOU, S. G. (2002). A Jump-Diffusion Model for Option Pricing. *Management Science* **48** 1086-1101.
- [14] KÜCHLER, U. and SØRENSEN, M. (1999). A note on limit theorems for multivariate martingales. *Bernoulli* **5** 483–493.
- [15] LOUKIANOVA, D. and LOUKIANOV, O. (2005). Uniform law of large numbers and consistency of estimators for Harris diffusions. *Statist. Probab. Lett.* **74** 347–355.
- [16] MAI, H. (2014). Efficient maximum likelihood estimation for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli* **20** 919–957.
- [17] MANCINI, C. (2011). The speed of convergence of the threshold estimator of integrated variance. *Stochastic Process. Appl.* **121** 845-855.
- [18] MASUDA, H. (2007). Ergodicity and exponential  $\beta$ -mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.* **117** 35-56.
- [19] MASUDA, H. (2009). Erratum to : "Ergodicity and exponential  $\beta$ -mixing bound for multidimensional diffusions with jumps". *Stochastic Processes and their Applications* **119** 676-678.
- [20] MASUDA, H. (2010). Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes. *Electronic Journal of Statistics* **4** 525-565.
- [21] MASUDA, H. (2013). Convergence of Gaussian quasi-likelihood random fields for ergodic Lévy driven SDE observed at high frequency. *Ann. Statist.* **41** 1593–1641.
- [22] MERTON, R. (1976). Option pricing when underlying stock returns are discontinuous. *J. Financial Economics* **3** 125–144.
- [23] OGIHARA, T. and YOSHIDA, N. (2011). Quasi-likelihood analysis for the stochastic differential equation with jumps. *Statistical Inference for Stochastic Processes* **14** 189-229.
- [24] REVUZ, D. . and YOR, M. (2005). *Continuous Martingales and Brownian Motion. (1997)*. Springer.
- [25] SHIMIZU, Y. (2006). M-Estimation for Discretely Observed Ergodic Diffusion Processes with Infinitely many Jumps. *Stat. Inference Stoch. Process.* **9** 179-225.
- [26] SHIMIZU, Y. (2008a). Some remarks on estimation of diffusion coefficient for jump-diffusions from finite samples. *Bull. Inform. Cybernet.* **40** 51-60.
- [27] SHIMIZU, Y. (2008b). A practical inference for discretely observed jump-diffusions from finite samples. *J. Japan. Statist. Soc.* **38** 391-413.
- [28] SHIMIZU, Y. and YOSHIDA, N. (2006). Estimation of Parameters for Diffusion Processes with Jumps from Discrete Observations. *Statistical Inference for Stochastic Processes* **9** 227-277.
- [29] TRAN, N. K. (2014). LAN property for jump diffusion processes with discrete observations via Malliavin calculus PhD thesis, University Paris 13.
- [30] VAN DER VAART, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic



Mathematics, Cambridge University Press.

- [31] YOSHIDA, N. (1992). Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis* **41** 220 - 242.

LAMME, UMR CNRS 8071  
UNIVERSITÉ D'ÉVRY VAL D'ESSONNE  
91037 ÉVRY CEDEX  
FRANCE  
E-MAIL: [arnaud.gloter@univ-evry.fr](mailto:arnaud.gloter@univ-evry.fr)  
[dasha.loukianova@univ-evry.fr](mailto:dasha.loukianova@univ-evry.fr)

CENTRE DE RECHERCHE EN ECONOMIE ET STATISTIQUE  
ENSAE-PARISTECH  
92245 MALAKOFF  
FRANCE  
E-MAIL: [hilmar.mai@ensae.fr](mailto:hilmar.mai@ensae.fr)