Série des Documents de Travail

n° 2016-04

Jump filtering and efficient drift estimation for Lévy-Driven SDE'S

A.Gloter[1]
D.Loukianova[2]
H.Mai[3]

[1] Université d'Évry Val d'Essonne. E-mail : arnaud.gloter@univ-evry.fr

[2] Université d'Évry Val d'Essonne. E-mail : Dasha.Loukianova@maths.univ-evry.fr

[3] ENSAE-CREST. E-mail : Hilmar.mai@gmail.com

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

ARNAUD GLOTER, DASHA LOUKIANOVA AND HILMAR MAI

ABSTRACT. The problem of efficient drift estimation for a parametric class of solutions of stochastic differential equations with Lévy-type jumps is considered under discrete high-frequency observations with growing observation window. The main challenge in this estimation problem stems from the two very different sources of noise involved: continuous diffusion and jump component of the process. This is reflected by the appearance of the unobserved continuous martingale part in the likelihood function. In order to obtain a feasible and efficient drift estimator based on discrete observations a jump filtering technique is employed to obtain a nonparametric estimators of integrals with respect to the continuous part. We prove general convergence results for these nonparametric estimators that are essential in any estimation problem concerning the continuous part such as drift an volatility estimation. Based on an LAN result for the general model this enables us finally to prove asymptotic efficiency in the sense of Hájek-Le Cam for the resulting drift estimator with jump filter. We then illustrate consequences of this general theory for a number of specific jump diffusion models, including the Cox-Ingersoll-Ross model with jumps from finance or the class of Ornstein-Uhlenbeck type processes. Another advantage of our approach are the straightforward implementation and the low computational costs which are demonstrated in a short simulation study that shows excellent agreement with our theoretical results.

1. INTRODUCTION

The class of solutions of Lévy-driven stochastic differential equations (SDEs) has recently attracted a lot of attention in the literature due to its many applications in various area such as a finance, physics and neuroscience. Indeed, it includes important examples from finance such as the well-known Barndorff-Nielsen-Shephard model, the Kou model and the Merton model (cf. Barndorff-Nielsen and Shephard [2001], Kou [2002] and Merton [1976]) as well as the stochastic Morris-Lecar neuron model (cf. for example Ditlevsen and Greenwood [2013]) from neuroscience to name just a few. Consequently, statistical inference for these models has become a recent and 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 $X^\theta_{t_0}, \dots, X^\theta_{t_n}$ of a jump diffusion process X^θ given by

$$X^\theta_t = X^\theta_0 + \int_0^t b(\theta, X^\theta_s)\, ds + \int_0^t \sigma(X^\theta_s)\, dW_s + \int_0^t \gamma(X^\theta_{s-})\, dL_s, \quad t \in \mathbb{R}_+, \tag{1}$$

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

We consider here the setting of high frequency observations with growing time window, i.e. for the discrete sample $X^\theta_{t_0}, \dots, X^\theta_{t_n}$ with $0 \leq t_0 \leq \ldots \leq t_n$ we assume that the sampling step

Key words and phrases. maximum likelihood estimation, jump diffusion process, Lévy process, efficient drift estimation, ergodic properties, jump filtering.

$\Delta_n := max\{t_i - t_{i-1} : 1 \leq i \leq n\}$ tends to 0 and $t_n \to \infty$ as $n \to \infty$. It is well known that due to the presence of the diffusion part one can only estimate the drift consistently if $t_n \to \infty$. From the point of view of applications, a crucial point in the high frequency setting is now to impose minimal conditions on the sampling step size Δ_n. This will be one of our main objectives in this work.

The topic of high frequency estimation for discretely observed diffusions without jumps is well developed by now. See for example Yoshida [1992] for joint estimation of drift and diffusion coefficient and Kessler [1997] for a Gaussian approximation of the transition density and the references therein. On the contrary, less results are known when a jump component is added to the problem. In the case of high frequency drift estimation for discretely observed diffusion with an additional jump component Masuda [2013] investigates Gaussian quasi-likelihood estimators of a joint drift-diffusion-jump part parameter. Shimizu and Yoshida [2006] define a contrast-type estimation function, for again joint estimation of drift, diffusion and jump part when the jumps are of compound Poisson type. Shimizu [2006] generalizes these results to include more general driving Lévy processes. The LAN property for drift and diffusion parameters is studied in Tran [2014] via Malliavin calculus techniques. In all these works 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 $n\Delta_n^2 \to 0$.

It is important to note here that the principles them self of the estimation of the drift, diffusion or jump law parameter 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 require 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. In this work we focus therefore on the estimation of the drift only and construct a consistent, asymptotically normal and efficient estimator, under minimal conditions on the jump behavior and the sampling scheme, which, in the case of the bounded jump measure density reduce to $n\Delta_n^{3-\varepsilon} \to 0$ for any $\epsilon > 0$.

Let us describe the outline of our method. Maximum likelihood estimation based on the likelihood function of the discrete sample is not feasible in this setting, since the likelihood function based on discrete observations 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 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 separation of the continuous and the jump part of an Itô-semimartingale appears naturally in many statistical inference questions (cf. for example Mancini [2011] and Bibinger and Winkelmann [2015]) and constitutes in itself an interesting nonparametric problem. In this article we study the question of nonparametric estimation of stochastic integrals with respect to the continuous part of X from a discrete sample of X. Proposition 6 and 71 give explicit rates of convergence for our estimators of these quantities. Besides being of independent interest these results constitute the main tool for the asymptotics analysis of our drift estimators.

The techniques 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 v_n, suggested by the typical behavior of a diffusion path. This approach will be called jump filtering in the sequel. Similar ideas of thresholding were also used in Shimizu and Yoshida [2006], Mancini [2011], Mai [2014] and Bibinger and Winkelmann [2015]. In this article we have payed particular attention to the study of the joint law of the biggest jump and of the total contribution of the other jumps in each sampling interval (Lemma 16), which permits us to improve existing conditions on the sampling scheme in the drift estimation problem.

A feasible 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 focus first on the MLE obtained from continuous observations and show that this MLE is asymptotically normal (Theorem 13) with explicit asymptotic variance. We then prove the LAN property which gives by Hàjek-Le Cam's convolution theorem that the continuous MLE is efficient (Theorem 14). 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 (Theorems 3, 4). The last step is mainly based on our results for the jump filter from Proposition 6 and 7.

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 Δ_n goes to 0, which depend on the behavior of the Lévy measure ν near zero. In the case where ν has a bounded Lebesgue density, these conditions reduce to $n\Delta_n^{3-\varepsilon} \to 0$ 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 10). This condition is in accordance with the results of Florens-Zimrou [1989] in the case of drift estimation for continuous diffusions, hence our result can be seen as a generalization of Florens-Zimrou [1989] to the presence of jumps.

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 prove the convergence of the jump filter. Section 5 is devoted to applications to popular parametric jump diffusion models and some numerical examples. Finally, in Section 6 and 7 we prove the main results and the convergence of the jump filter, respectively, and Section 8 contains some auxiliary results that are frequently used in the sequel.

2. Model, assumptions and ergodicity

Let Θ be a compact subset of $\mathbb{R}^d$ and X^θ a solution to (1) which can be rewritten as

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

where $W = (W_t)_{t\geq 0}$ is a one-dimensional Brownian motion and μ 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 $\int_{\mathbb{R}} |z| d\nu(z) < \infty$. The initial condition X_0^θ, W and L are independent. We assume without loss of generality that $0 \in \Theta$ and $b(0, \cdot) \equiv 0$.

2.1. Assumptions.

We suppose that the functions $b : \Theta \times \mathbb{R} \to \mathbb{R}$, $\sigma : \mathbb{R} \to \mathbb{R}$ and $\gamma : \mathbb{R} \to \mathbb{R}$ 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 Θ.*

Under Assumption 1 equation (1) admits a unique non-explosive càdlàg adapted solution possessing the strong Markov property, cf. Applebaum [2009](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^θ 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 $z = z(x, U) \in supp(\nu)$ such that $\gamma(x)z \in U$.

The last Assumption was used in Masuda [2007] to prove the irreducibility of the process X^θ. See also Masuda [2009] 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 $|x| \to \infty$.*
(iii): *$|\gamma(x)|/|x| \to 0$ as $|x| \to \infty$.*
(iv): *$|\sigma(x)|/|x| \to 0$ as $|x| \to \infty$.*
(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 π^θ, as well as the ergodicity of the process X^θ, as stated in Lemma 1 below.

Assumption 4 (Jumps). **(i):** *The jump coefficient γ is bounded from below, i.e. $\inf_{x\in\mathbb{R}} |\gamma(x)| := \gamma_{min} > 0$ (wlog we suppose $\gamma_{min} \geq 1$).*
(ii): *the Lévy measure ν satisfies $\int_{0<|z|\leq 1} |z|\nu(dz) < \infty$,*
(iii): *the Lévy measure ν is absolutely continuous with respect to the Lebesgue measure,*
(iv): *the jump coefficient γ is upper bounded, i.e. $\sup_{x\in\mathbb{R}} |\gamma(x)| := \gamma_{max} < \infty$.*

Note that the integrability condition given by the Assumption 4 (ii) is automatically satisfied in the finite activity case $\nu(\mathbb{R}) < \infty$. This condition 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) can be defined as a deterministic Lebesgue-Stieltjes integral. The third and the fourth point of the Assumption 4 are need in the infinite activity case.

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 17) that this last assumption is equivalent to

$$\forall \theta \neq \theta', \quad (\theta, \theta') \in \Theta^2, \quad b(\theta, .) \neq b(\theta', .). \tag{2}$$

For $f : \Theta \to \mathbb{R}$ denote by $\nabla_\theta f : \Theta \to \mathbb{R}^d$ the gradient column vector and by $\partial_\theta^2 f := \left(\partial^2_{\theta_i,\theta_j} f\right)_{1\leq i,j\leq d}$ the Hessian matrix of f. We define $|\theta|$ as the Euclidian norm of $\theta \in \mathbb{R}^d$, and $|\partial_\theta^2 f| := \sqrt{\sum_{i,j=1}^n |\partial^2_{\theta_i,\theta_j} f|^2}$ as the Euclidian norm of the Hessian matrix of f. The following assumption is used to insure the uniform in θ convergence needed in the proofs of consistency and asymptotic normality:

Assumption 7 (Hölder-continuity of drift). **(i):** *For all $x \in \mathbb{R}$, $b(., x)$ 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 $K : \mathbb{R} \to \mathbb{R}_+$ is at most of polynomial growth.

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

$$\forall \theta, \theta', \quad |\nabla b(\theta,x) - \nabla b(\theta',x)| \le K_1(x)|\theta-\theta'|^{\kappa_1}$$
$$\forall \theta, \theta', \quad |\partial_\theta^2 b(\theta,x) - \partial_\theta^2 b(\theta',x)| \le K_2(x)|\theta-\theta'|^{\kappa_2}$$

where $0 < \kappa_1, \kappa_2 \le 1$ and $K_1, K_2 : \mathbb{R} \to \mathbb{R}_+$ are at most of polynomial growth.

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 σ', σ'' as well as the functions*

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

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

Define the asymptotic Fisher information by

$$I(\theta) = \left(\int_{\mathbb{R}} \frac{\partial_{\theta_i} b(\theta,x) \partial_{\theta_j} b(\theta,x)}{\sigma^2(x)} \pi^\theta(dx) \right)_{1 \le i,j \le d}. \tag{3}$$

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). The following lemma is a generalization of a result of Masuda [2007]. It states conditions for the existence of an invariant measure π^θ such that an ergodic theorem holds and moments of all order exist. A proof is given in Section 8.

Lemma 1. *Under assumptions (1) to (4), for all $\theta \in \Theta$, X^θ admits a unique invariant distribution π^θ and the ergodic theorem holds:*

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

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

(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,*

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

3. Construction of the estimator and main results

We define a discrete approximation to 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^θ be given by (1). We denote by P^θ the law of X^θ on the Skorokhod space $D[0,\infty)$ of real-valued càd làg functions, and P^θ_t its restriction on $D[0,t)$. From now on we denote the true parameter value by $\theta^\star$, an interior point of the parameter space Θ that we want to estimate. We shorten X for $X^{\theta^\star}$ and P, E, π for respectively $P^{\theta^\star}, E^{\theta^\star}, \pi^{\theta^\star}$. Suppose that we observe a finite sample

$$X_{t_0}, \dots, X_{t_n}; \quad 0 = t_0 \le t_1 \le \dots \le t_n. \tag{4}$$

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.

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

We assume $\lim_{n\to\infty} t_n = \infty$ and $n\Delta_n = O(t_n)$ as $n \to \infty$. Under Assumption 5, P^θ_t and P^0_t are mutually locally absolutely continuous for any $\theta \in \Theta$ (cf. for example Jacod and Shiryaev [2003]) and the likelihood function is given by

$$\mathcal{L}_t(\theta, X) = \frac{dP^\theta_t}{dP^0_t}(X) = \exp\left(\int_0^t \sigma(X_s)^{-2} b(\theta, X_s)\, dX^c_s - \frac{1}{2}\int_0^t \sigma(X_s)^{-2} b(\theta, X_s)^2\, ds\right). \tag{5}$$

We define the log-likelihood function as

$$\ell_t(\theta) := \ln \mathcal{L}_t(\theta, X). \tag{6}$$

The crucial point here is the appearance of X^c in (5), since when X is observed discretely, its continuous part remains unknown. To handle this problem we use a jump filter as described below.

For $g : [0, t_n] \to \mathbb{R}$, set $\Delta^n_i g = g_{t_i} - g_{t_{i-1}}$, $i = 1, \dots n$. In particular, $\Delta^n_i X = X_{t_i} - X_{t_{i-1}}$, $\Delta^n_i X^c = X^c_{t_i} - X^c_{t_{i-1}}$ and $\Delta^n_i Id = t_i - t_{i-1}$. Let $\varepsilon \in (0, 1/2)$ and denote

$$v_n = \Delta_n^{1/2-\varepsilon},\ n \ge 1. \tag{7}$$

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

$$\ell^n_{t_n}(\theta) = \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} b(\theta, X_{t_{i-1}}) \Delta^n_i X \mathbf{1}_{|\Delta^n_i X| \le v_n} - \frac{1}{2}\sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} b(\theta, X_{t_{i-1}})^2 \Delta^n_i Id. \tag{8}$$

The cut-off sequence (v_n) 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 (7). The challenge is now to find suitable conditions on Δ_n, ϵ and ν is the likelihood in (6) well approximated by its discretized and jump filtered counterpart (8) even in the case of infinite activity. Of course we can choose ε arbitrarily small, which is a choice we have in mind. Finally, we define an estimator $\hat\theta_n$ of $\theta^\star$ as

$$\hat\theta_n \in \operatorname*{argmax}_{\theta\in\Theta} \ell^n_{t_n}(\theta) \tag{9}$$

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, Δ_n and v_n.

Theorem 2 (Consistency)**.** *Suppose that Assumptions 1 to 8 hold, then the FMLE $\hat\theta_n$ is consistent in probability:*

$$\hat\theta_n \xrightarrow{P} \theta^\star, \qquad n \to \infty.$$

To obtain a central limit theorem for the estimation error we consider finite and infinite activity separately, since we obtain different conditions on the relation of n, Δ_n and the cut-off sequence v_n.

Theorem 3 (Asymptotic normality: finite activity)**.** *Assume that the Lévy process L has a finite jump activity : $\nu(\mathbb{R}) < \infty$. Suppose that Assumptions 1 to 3, 4(i) and 6 to 9 hold.*

If $n\Delta_n^{3-\varepsilon} \to 0$, $\sqrt{n}\Delta_n^{1-\varepsilon/2}\left(\int_{|z|\le 2v_n}\nu(dz)\right)^{1-\varepsilon/2} \to 0$ and $\sqrt{n}\Delta_n^{1/2}\int_{|z|<2v_n}|z|\nu(dz) \to 0$ as $n \to \infty$, then we conclude that the FMLE $\hat{\theta}_n$ is asymptotically normal:

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

where I is the Fisher information given by (3).

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

Remark 1. If ν has a bounded Lebesgue density, the conditions of the Theorem 3 on the sampling scheme and the jump behavior reduce to $n\Delta_n^{3-4\varepsilon} \to 0$.

The following theorem generalizes the results of Theorem 3 to driving Lévy processes of infinite activity.

Theorem 4 (Asymptotic normality: general case)**.** *Assume that the Lévy process L has infinite jump activity : $\nu(\mathbb{R}) = \infty$. Suppose Assumptions 1 to 9 hold. If $n\Delta_n^{3-\varepsilon} \to 0$,*

$$\sqrt{n\Delta_n}\left(\int_{|z|\le 3v_n/\gamma_{min}}|z|\nu(dz)\right)^{1-\varepsilon/2} \to 0 \quad \text{and} \quad \sqrt{n}\Delta_n^{3/2-2\varepsilon}\left(\int_{|z|\ge 3v_n/\gamma_{min}}\nu(dz)\right)^{1-\varepsilon/2} \to 0$$

as $n \to \infty$, then all conclusions of Theorem 3 hold.

Theorem 4 applies for both finite and infinite jump activity. Besides different conditions on the sampling scheme and the behavior of ν near zero it uses that the Lévy measure ν admits a density, which is not supposed in Theorem 3. In the case where ν admits a bounded Lebesgue density, all the conditions on the Δ_n and n of the Theorem 4 reduce to $n\Delta_n^{3-\tilde{\varepsilon}} \to 0$ for some $\tilde{\varepsilon} > 0$ as in the Theorem 3.

Example 5 (tempered stable jumps)**.** *To illustrate the influence of the jump behavior of L on the conditions on n and Δ_n given in Theorem 4 let us consider the example of a tempered α-stable driving Lévy process. Tempered stable processes have been popular in financial modeling to overcome the limitations of the classical models based on Brownian motion alone (cf. Cont and Tankov [2004]). The Lévy measure in this case has an unbounded and non-integrable density given by*

$$\nu(dz) = C|z|^{-(1+\alpha)}e^{-\lambda|z|}dz$$

with $\lambda > 0$ and a normalizing constant $C > 0$ that satisfies the conditions of Theorem 4 if $0 < \alpha < 1$.

The conditions on n, Δ_n and ν in Theorem 4 can now be summarized as $n\Delta_n^{2-\alpha-\tilde{\epsilon}} \to 0$ for some $\epsilon > 0$. We observe that a higher Blumenthal-Getoor index α requires a faster convergence Δ_n to zero. This is in line with the intuition that when the intensity of small jumps increases (i.e. α 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. The following proposition concerns the finite activity case. The cut-off sequence v_n and ε were defined in (7).

Proposition 6 (jump filtering: finite activity). *Suppose that L is of finite activity and Assumptions 1 to 4 hold. Suppose that $f : \Theta \times \mathbb{R} \to \mathbb{R}$ satisfies:*

a) *for all $x \in \mathbb{R}$, $f(.,x)$ is Hölder continuous with respect to $\theta \in \Theta$:*
$$\forall \theta, \theta', \quad |f(\theta,x) - f(\theta',x)| \leq C(x)|\theta-\theta'|^{\kappa},$$
where $0 < \kappa \leq 1$ and $C : \mathbb{R} \to \mathbb{R}_+$ is at most of polynomial growth;

b) *for all $\theta \in \Theta$, $f(\theta,.) \in \mathcal{C}^2(\mathbb{R})$ and $\sup_{\theta\in\Theta}|f(\theta,.)|$, $\sup_{\theta\in\Theta}|f'_x(\theta,.)|$ and $\sup_{\theta\in\Theta}|f''_x(\theta,.)|$ are at most of polynomial growth.*

Then the following statements hold:

(i) *without any assumption on the way that $\Delta_n \to 0$ as $n \to \infty$,*
$$(n\Delta_n)^{-1} \sup_{\theta\in\Theta} \left| \int_0^{t_n} f(\theta, X_s)\, dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n} \right| \xrightarrow{P} 0;$$

(ii) *if $n\Delta_n^{3-\varepsilon} \to 0$, $\sqrt{n}\Delta_n^{1-\varepsilon/2} \left(\int_{|z|\leq 2v_n} \nu(dz)\right)^{1-\varepsilon/2} \to 0$ and $\sqrt{n}\Delta_n^{1/2} \int_{|z|\leq 2v_n} |z|\nu(dz) \to 0$ as $n \to \infty$, then for any $\theta \in \Theta$,*

$$(n\Delta_n)^{-1/2} \left| \int_0^{t_n} f(\theta, X_s)\, dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n} \right| \xrightarrow{P} 0. \tag{10}$$

The case of infinite activity is treated in the following proposition.

Proposition 7 (jump filtering: infinite activity). *Suppose that L is of infinite activity and Assumptions 1 to 4 hold. Suppose that $f : \Theta \times \mathbb{R} \to \mathbb{R}$ satisfies the assumptions of Proposition 6. Then,*

(i) *statement (i) of Proposition 6 holds;*

(ii) *if $n\Delta_n^{3-\varepsilon} \to 0$,*
$$\sqrt{n\Delta_n}\left(\int_{|z|\leq 3v_n/\gamma_{min}} |z|\nu(dz)\right)^{1-\varepsilon/2} \to 0 \quad \text{and} \quad \sqrt{n}\Delta_n^{3/2-\varepsilon}\left(\int_{|z|\geq 3v_n/\gamma_{min}} \nu(dz)\right)^{1-\varepsilon/2} \to 0$$
as $n \to \infty$, then for any $\theta \in \Theta$, the convergence (10) holds.

The proofs of both propositions are based on the following three lemmas. Lemma 8 and 9 describe the approximation of the discretized stochastic integral with respect to X^c by the jump filter in the cases of finite and infinite activity, respectively. To prove the propositions 6 and 7 we also need a convergence result for the Euler scheme in order to approximate the stochastic integral with respect to X^c by the corresponding discrete sum. This will be done in Lemma 10.

Lemma 8 (jump filtering error: finite activity)**.** *Assume that L is of finite activity and $f:\Theta\times\mathbb{R}\to\mathbb{R}$ is such that $\sup_{\theta\in\Theta}|f(\theta,x)|$ is sub-polynomial. Under Assumption 1 to 4, we obtain*

(i)
$$\sup_{\theta\in\Theta}|\sum_{i=1}^n f(\theta,X_{t_{i-1}})\left(\Delta_i^n X^c-\Delta_i^n X\mathbf{1}_{|\Delta_i^n X|\le v_n}\right)|=O_{\mathbb{L}^1}(n\Delta_n^{3/2-\varepsilon/2}).$$

(ii) *for all $\theta\in\Theta$, if $n\Delta_n^{3-\varepsilon}\to 0$ as $n\to\infty$,*
$$\sum_{i=1}^n f(\theta,X_{t_{i-1}})\left(\Delta_i^n X^c-\Delta_i^n X\mathbf{1}_{|\Delta_i^n X|\le v_n}\right)=o_P\left(\sqrt{n\Delta_n}\right)$$
$$+O_{\mathbb{L}^1}\left(n\Delta_n^{5/2-\varepsilon}+n\Delta_n^{3/2-\varepsilon/2}\left(\int_{|z|\le 2v_n}\nu(dz)\right)^{1-\varepsilon/2}+n\Delta_n\int_{|z|\le 2v_n}|z|\nu(dz)\right).$$

The next lemma extends the uniform bound to the case of infinite activity.

Lemma 9 (jump filtering error: infinite activity)**.** *Assume that L is of infinite activity and $f:\Theta\times\mathbb{R}\to\mathbb{R}$ is such that $\sup_{\theta\in\Theta}|f(\theta,x)|$ is sub-polynomial.*

(i) *Under Assumption 1 to 4, we obtain*
$$\sup_{\theta\in\Theta}|\sum_{i=1}^n f(\theta,X_{t_{i-1}})\left(\Delta_i^n X^c-\Delta_i^n X\mathbf{1}_{|\Delta_i^n X|\le v_n}\right)|=$$
$$O_{\mathbb{L}^1}\left(n\Delta_n\left(\int_{|z|\le 3v_n}|z|\nu(dz)\right)^{1-\varepsilon/2}+n\Delta_n^{3/2-\varepsilon}\left(\int_{|z|\ge v_n/\gamma_{min}}\nu(dz)\right)^{1-\varepsilon/2}\right)$$

(ii) *for all $\theta\in\Theta$, if $n\Delta_n^{3-\varepsilon}\left(\int_{|z|\ge 3v_n/\gamma_{min}}\nu(dz)\right)^{2-\varepsilon}\to 0$, as $n\to\infty$, then*
$$\sum_{i=1}^n f(\theta,X_{t_{i-1}})\left(\Delta_i^n X^c-\Delta_i^n X\mathbf{1}_{|\Delta_i^n X|\le v_n}\right)=o_P\left(\sqrt{n\Delta_n}\right)$$
$$+o_{\mathbb{L}^1}\left(n\Delta_n^{2-\varepsilon}(\int_{|z|\ge 3v_n/\gamma_{min}}\nu(dz))^{1-\varepsilon/2}\right)+O_{\mathbb{L}^1}\left(n\Delta_n(\int_{|z|\le 3v_n/\gamma_{min}}|z|\nu(dz))^{1-\varepsilon/2}\right).$$

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

Lemma 10 (Euler scheme)**.** *Suppose that $f:\Theta\times\mathbb{R}\to\mathbb{R}$ satisfies the following assumptions:*

a) *for all $x\in\mathbb{R}$, $f(.,x)$ is Hölder continuous with respect to $\theta\in\Theta$:*
$$\forall\theta,\theta',\quad |f(\theta,x)-f(\theta',x)|\le K(x)|\theta-\theta'|^\kappa;$$
where $0<\kappa\le 1$ and $K:\mathbb{R}\to\mathbb{R}_+$ is at most of polynomial growth;

b) *for all $\theta\in\Theta$, $f(\theta,.)\in\mathcal{C}^2(\mathbb{R})$ and $\sup_{\theta\in\Theta}|f(\theta,.)|$, $\sup_{\theta\in\Theta}|f'_x(\theta,.)|$ and $\sup_{\theta\in\Theta}|f''_x(\theta,.)|$ are at most of polynomial growth.*

Under Assumptions 1 to 4, we obtain

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

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

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

We have now collected all the tools to prove the convergence of the jump filter approximation towards integral functionals with respect to the continuous martingale part as stated in Proposition 6 and 7.

Proof of Proposition 6. We decompose the difference as follows:

$$\left| \int_0^{t_n} f(\theta, X_s)\, dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \le v_n} \right| \le \tag{11}$$

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

We first prove (i). By Lemma 10, the first term on the right hand side of (11) divided by $n\Delta_n$ goes to zero uniformly, without any condition on Δ_n. Combining it with (i) of the Lemma 8 we get the result.

We now prove (ii). For the first term of (11) divided by $(n\Delta_n)^{1/2}$ we use (ii) of the Lemma 10. Moreover, the (ii) of the Lemma 8, gives, for any $\theta \in \Theta$,

$$(n\Delta_n)^{-1/2} \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \le v_n} \right) = o_P(1) +$$

$$O_{\mathbb{L}^1} \left(\sqrt{n}\Delta_n^{2-\varepsilon} + \sqrt{n}\Delta_n^{1-\varepsilon/2} \left(\int_{|z| \le 2v_n} \nu(dz) \right)^{1-\varepsilon/2} + \sqrt{n}\Delta_n^{1/2} \int_{|z| \le 2v_n} |z| \nu(dz) \right) \xrightarrow{P} 0$$

under conditions (ii) of the proposition.

□

Proof of Proposition 7. We use the decomposition (11)

and prove first the statement (i). Using the Lemma 10 the first term of (11) divided by $n\Delta_n$ goes to zero uniformly without any condition on Δ_n.

Lemma 9 together with the Assumption 4 (ii) and the fact that $v_n = \Delta_n^{1/2-\varepsilon}$ gives

$$(n\Delta_n)^{-1} \sup_\theta | \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i X| \le v_n} \right) | =$$

$$O_{\mathbb{L}^1} \left(\left(\int_{|z| \le 3v_n} |z| \nu(dz) \right)^{1-\varepsilon/2} + \Delta_n^{1/2-\varepsilon/2} \left(\int_{|z| \ge v_n/\gamma_{min}} \nu(dz) \right)^{1-\varepsilon/2} \right) =$$

$$O_{\mathbb{L}^1} \left(\left(\int_{|z| \le 3v_n} |z| \nu(dz) \right)^{1-\varepsilon/2} + \frac{\Delta_n^{1/2-\varepsilon/2}}{v_n^{1-\varepsilon/2}} \left(\int_{|z| \ge v_n/\gamma_{min}} |z| \nu(dz) \right)^{1-\varepsilon/2} \right) \xrightarrow{P} 0.$$

Hence statement (i) is proved.

Now we prove statement (ii). For any $\theta \in \Theta$, under the condition $n\Delta_n^{3-\varepsilon} \to 0$, the second statement of Lemma 10 gives the convergence to 0 of the first term in the decomposition (11), divided by

$\sqrt{n}\Delta_n$. The convergence to 0 of the second term of (11), divided by $\sqrt{n}\Delta_n$, immediately follows from Lemma 9 and the conditions of (ii). □

When discretizing the likelihood function, we need the following lemma, whose proof can be found in the Section 8.

Lemma 11. *Suppose that Assumptions 1–4 are satisfied . Suppose that $f : \Theta \times \mathbb{R} \to \mathbb{R}$ is such that $\forall \theta \in \Theta$, $f(\theta, .) \in \mathcal{C}^1(\mathbb{R})$ and $\sup_{\theta\in\Theta} |f'(\theta, .)|$ is sub-polynomial. Then we obtain:*

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

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

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

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

5. Examples and numerical results

In this section we consider concrete applications of the drift estimator in popular jump diffusion models and investigate the numerical performance in finite sample studies. We consider both examples with finite and infinite jump activity.

In the first part we give explicit drift estimators for Ornstein-Uhlenbeck-type and CIR processes and compare there performance in a Monte Carlo study for finite activity jumps. Then we apply our method to a hyperbolic diffusion process with α-stable jump component of infinite jump activity. We consider here for convenience only linear models in the drift parameter that lead to explicit maximum likelihood estimators in order to avoid the need for numerical maximization techniques. Note that the method developed in this work applies equally well to non-linear models by using standard maximization methods on the discretized and jump-filtered likelihood function (8).

It turns out that our estimators can be applied even beyond the scope of our theoretical results. To demonstrate this we include in Section 5.2 models that do not posses moments of all orders and consider Lévy processes of unbounded variation in our simulations.

5.1. Finite activity.

In this section we consider two different jump diffusion models with finite activity jumps. The first model will consist of Ornstein-Uhlenbeck-type processes that recently became popular in financial modeling (cf. for example Barndorff-Nielsen and Shephard [2001]). In the second part we extend a Cox-Ingersoll-Ross model from finance (cf. Cox et al. [1985]) by including jumps and investigate the finite sample behavior of the drift estimator and jump filter for varying observation settings. The jump process L is of compound Poisson type in the case of finite activity such that it can be written as

$$L_t = \sum_{i=1}^{N_t} Z_i, \text{ for } t \geq 0, \tag{12}$$

where $(N_t)_{t\geq 0}$ is a Poisson process with intensity λ and $(Z_i)_{i\in\mathbb{N}}$ are i.i.d. real random variables independent of N, with distribution ν/λ.

5.1.1. *Ornstein-Uhlenbeck-type processes.* Suppose that we have given a discrete sample

$$X_{t_0}, \dots, X_{t_n} \quad \text{for } t_i = i\Delta_n \text{ and } i = 0, \dots, n, \tag{13}$$

of an Ornstein-Uhlenbeck-type (OU) process $(X_t)_{t\geq 0}$ that is defined as a solution of the stochastic differential equation

$$dX_t = (\theta_2 - \theta_1 X_t)\, dt + \sigma\, dW_t + dL_t \quad X_0 = x,$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion and $(L_t)_{t\geq 0}$ a pure jump Lévy process. Our goal is to estimate the unknown drift parameter $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$. The volatility parameter $\sigma > 0$ might be unknown and can be seen as a nuisance parameter. The jump component $(L_t)_{t\geq 0}$ will be of compound Poisson type, i.e. it can be written as in (12) with intensity λ and the jump heights Z_i are supposed to be iid with exponential distribution with rate 1.

From (8) and (9) we find that the FMLE for θ is the solution $\hat\theta_n^{\mathrm{OU}} = (\hat\theta_{1,n}^{\mathrm{OU}}, \hat\theta_{2,n}^{\mathrm{OU}})$ to the following set of linear equations in θ_1 and θ_2.

$$\begin{aligned}\theta_1 &= \frac{\theta_2 I_n(X,1) - \sum_{i=1}^n X_{t_i} \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n}}{I_n(X^2)},\\ \theta_2 &= \frac{\sum_{i=1}^n \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n} + \theta_1 I_n(X,1)}{t_n},\end{aligned} \tag{14}$$

where we introduced the functional

$$I_n(X,p) := \sum_{i=1}^n X_{t_i}^p \Delta_i^n Id \quad \text{for } p \in \mathbb{R}. \tag{15}$$

The FLME for the first component of θ results in

$$\hat\theta_{1,n}^{\mathrm{OU}} = \left(1 - \frac{I_n(X,1)^2}{I_n(X,2)}\right)^{-1} \frac{I_n(X,1)\sum_{i=1}^n \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n} - t_n \sum_{i=1}^n X_{t_i} \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n}}{t_n I_n(X,2)}.$$

The second component $\hat\theta_{2,n}^{\mathrm{OU}}$ follows now easily by plugging $\hat\theta_{1,n}^{\mathrm{OU}}$ into (14).

In Table 1 we give simulation results for $\hat\theta_{1,n}^{\mathrm{OU}}$. The given mean and standard deviation are each based on 500 Monte Carlo samples of $\hat\theta_{1,n}^{\mathrm{OU}}$. In this example we choose $v_n = \Delta_n^{1/3}$ in order to approximate well the continuous martingale part that appeared in the likelihood function (5). We compare different observation schemes and different jump intensities λ for true parameter values given by $\theta_1 = 2$ and $\theta_2 = 0$. The drift estimator performs well over the whole range of settings provide that the discretization distance Δ_n is sufficiently small. We also give the average number of jumps that were detected by the jump filter and observe that this number scales as expected linearly in t_n.

5.1.2. *Cox-Ingersoll-Ross (CIR) processes with jumps.* We define a CIR or square-root process $X = (X_t)_{t\geq 0}$ with jumps as a solution to the SDE

$$X_t = (\theta_1 - \theta_2 X_t)\, dt + \sigma\sqrt{X_t}\, dW_t + dL_t,$$

where $\theta_1, \theta_2, \sigma > 0$, (W_t) is a standard Brownian motion and (L_t) a pure jump Lévy process. The two-dimensional drift parameter $\theta = (\theta_1, \theta_2)$ is unknown and will be estimated from discrete observations of X as in (13).

The classical CIR process without jumps (e.g. $L_t \equiv 0$) has the property that it stays non-negative at all times which makes it an interesting model for financial applications e.g. in interest rate modeling (Vasicek model) and stochastic volatility models (Heston model). We consider here

		$\lambda = 1$			$\lambda = 6$		
t_n	n	mean	std dev	jumps filt	mean	std dev	jumps filt
2	100	1.4	0.7	6.5	1.4	0.6	15.8
	300	1.8	0.8	6.8	1.7	0.6	15.9
	600	2.0	0.8	7.9	1.9	0.5	16.3
	800	2.0	0.8	7.2	2.0	0.6	16.5
5	600	1.4	0.6	13.1	1.3	0.39	39.5
	1200	1.8	0.6	13.6	1.7	0.39	40.4
	4000	2.0	0.7	13.6	1.8	0.39	41.4
	6000	2.1	0.7	12.4	1.9	0.37	41.5
10	600	1.2	0.26	19.1	1.3	0.21	67
	2000	2.0	0.27	21.6	1.6	0.2	75

TABLE 1. Monte Carlo estimates of mean and standard deviation from 500 samples of $\hat{\theta}^{\mathrm{OU}}_{1,n}$ for an OU process with compound Poisson jumps with intensity λ and true parameter $\theta_1 = 2$.

therefore a jump component (L_t) of compound Poisson type that exhibits only positive jumps such that X will stay non-negative. In fact, we take a driving Lévy process (L_t) as in (12) with intensity $\lambda = 1$ and exponentially distributed jumps with rate $\eta > 0$, e.g. $Z_i \sim \mathrm{Exp}(\eta)$.

The filtered maximum likelihood estimator for θ is this model can be easily derived from (8) and (9). It is given as the solution $\hat{\theta}^{\mathrm{CIR}}_n = (\hat{\theta}^{\mathrm{CIR}}_{1,n}, \hat{\theta}^{\mathrm{CIR}}_{2,n})$ to the following set of linear equations in the parameters θ_1 and θ_2.

$$\theta_1 = \frac{\theta_2 t_n - \sum_{i=1}^n X_{t_i}^{-1}\Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \le v_n}}{I_n(X,-1)}, \quad \theta_2 = \frac{\theta_1 t_n - \sum_{i=1}^n \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \le v_n}}{I_n(X,1)}, \tag{16}$$

where $I_n(X,p)$ for $p \in \mathbb{R}$ was defined in (15). We obtain for $\hat{\theta}^{\mathrm{CIR}}_{2,n}$ the FMLE

$$\hat{\theta}^{\mathrm{CIR}}_{2,n} = (I_n(X,-1)I_n(X,1) - t_n)^{-1}\left(\sum_{i=1}^n X_{t_i}^{-1}\Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \le v_n} - I_n(X,-1)\sum_{i=1}^n \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \le v_n}\right).$$

The first component $\hat{\theta}^{\mathrm{CIR}}_{1,n}$ follows now immediately by plugging $\hat{\theta}^{\mathrm{CIR}}_{2,n}$ into (16).

To obtain Monte Carlo estimates of mean and standard deviation of $\hat{\theta}^{\mathrm{CIR}}_n$ we simulate discrete samples of X on an equidistant grid as in the previous example. We take $v_n = \Delta_n^{1/3}$ in order to approximate the continuous martingale part of X. In Table 2 we report the results for $\hat{\theta}^{\mathrm{CIR}}_{2,n}$ from 1000 Monte Carlo samples each. The results are given for different t_n, n and σ for true parameter values $\theta_1 = 0.1$ and $\theta_2 = 2$. We find that $\hat{\theta}^{\mathrm{CIR}}_{2,n}$ performs well as long as the discretization step size Δ_n is fine enough such that a high-frequency approximation becomes valid.

5.2. **Infinite activity.** In this section we investigate estimation of the drift when the driving Lévy process is of infinite jump activity. This is of course a more challenging problem with regards to the approximation of the continuous martingale part i.e. the jump filtering problem, since we have to distinguish a diffusion component from a process that jumps infinitely often in finite time intervals.

		$\sigma = 0.25$			$\sigma = 0.5$		
t_n	n	mean	std dev	jumps filt	mean	std dev	jumps filt
5	200	1.7	0.22	6.8	1.7	0.28	8.0
	400	1.9	0.12	5.1	1.8	0.2	6.6
	800	2.0	0.09	4.5	1.9	0.17	5.6
10	500	1.7	0.15	12	1.7	0.21	15
	1000	1.9	0.08	9.7	1.8	0.14	12
	1500	1.9	0.06	9.5	1.9	0.13	11
20	1000	1.8	0.13	25	1.6	0.16	30
	2000	1.9	0.06	19	1.8	0.11	24
	3000	2.0	0.04	19	1.9	0.09	22

TABLE 2. Monte Carlo estimates of mean and standard deviation of $\hat{\theta}_{2,n}^{\mathrm{CIR}}$ for a CIR process with Gaussian component and compound Poisson jumps with intensity $\lambda = 1$ and true drift parameter $\theta_2 = 2$.

5.2.1. *Hyperbolic diffusions with jumps.* In this section we apply the drift estimator to hyperbolic diffusion processes with jumps. They are defined as solutions $(X_t)_{t\geq 0}$ of the following SDE:

$$dX_t = -\frac{\theta X_t}{(1+X_t^2)^{1/2}}\,dt + \sigma\,dW_t + dL_t, \quad X_0 = x.$$

Here, the drift parameter $\theta > 0$ and the diffusion coefficient $\sigma > 0$ are unknown and we aim at estimating θ form discrete observations $X_{t_0}, \ldots, X_{t_n}$ of X, where $t_i = i\Delta_n$ for $\Delta_n > 0$ and $i = 0, \ldots, n$. The driving Lévy process $(L_t)_{t\geq 0}$ will be an α-stable process with Lévy-Khintchine triplet $(0, 0, \nu)$ such that the Lévy measure is of the form $\nu(dx) = dx/|x|^{1+\alpha}$.

From (5) we obtain an explicit pseudo MLE for θ in this model class given by

$$\hat{\theta}_t^{\mathrm{hyp}} = -\frac{\int_0^t \frac{X_s}{(1+X_s^2)^{1/2}}\,dX_s^c}{\int_0^t \frac{X_s^2}{(1+X_s^2)}\,ds}.$$

Via discretization and jump filtering this leads to the following drift estimator based on discrete observations:

$$\hat{\theta}_n^{\mathrm{hyp}} = -\sum_{i=1}^n \frac{X_{t_i}}{(1+X_{t_i}^2)^{1/2}} \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n} \left(\sum_{i=1}^n \frac{X_{t_i}^2}{(1+X_{t_i}^2)} \right)^{-1}$$

To assess the performance of $\hat{\theta}_n^{\mathrm{hyp}}$ in Monte Carlo experiments we simulate discrete trajectories of X via a Euler scheme with sufficient small step size.

In Table 3 we give estimated mean and standard deviation of $\hat{\theta}_n^{hyp}$ from 500 Monte Carlo samples each for different observation length t_n and number of observations n. We consider two different values for the index of stability α and give also the number of jumps that have been detected by the jump filter. It turns out that $\hat{\theta}_n^{hyp}$ performs remarkably well over the whole range of different setting even in the case $\alpha = 1$ of infinite variation jumps that is not covered by our theoretical results, since we have assumed that $\int_{\mathbb{R}} |x|\nu(dx) < \infty$. It might therefore be reasonable to expect that the convergence results presented here can be extended to jumps processes with Blumenthal-Getoor index $\alpha \geq 1$.

		$\alpha = 0.5$			$\alpha = 1$		
t_n	n	mean	std dev	jumps filt	mean	std dev	jumps filt
5	600	1.7	0.53	26	1.6	0.62	37
	1200	1.9	0.54	27	1.8	0.60	40
	1500	1.9	0.57	26	1.9	0.66	41
10	1000	1.6	0.33	51	1.5	0.40	71
	2000	1.8	0.34	53	1.7	0.38	79
	4000	1.9	0.35	50	1.9	0.43	85
20	2000	1.6	0.23	104	1.6	0.27	142
	4000	1.8	0.24	106	1.7	0.28	158
	8000	1.9	0.23	101	1.9	0.30	170

TABLE 3. Monte Carlo estimates of mean and standard deviation from 500 samples of $\hat{\theta}_n^{\mathrm{hyp}}$ for a hyperbolic diffusion process with Gaussian component and α-stable jumps and true drift parameter $\theta = 2$

6. Proofs of main results

6.1. **MLE for continuous observations.** Let $\bar{\theta}_t$ be the true MLE maximizing the log-likelihood function given by (6) and based on continuous observations :

$$\bar{\theta}_t \in \operatorname*{argmax}_{\theta \in \Theta} \ell_t(\theta). \tag{17}$$

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

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

Proof. Denote

$$\tilde{\ell}_t(\theta) := \int_0^t \frac{(b(\theta, X_s) - b(\theta^\star, X_s))}{\sigma(X_s)} dW_s - \frac{1}{2}\int_0^t \frac{(b(\theta, X_s) - b(\theta^\star, X_s))^2}{\sigma^2(X_s)} ds. \tag{18}$$

Using (1) and the fact that the observed trajectory corresponds to the true value of parameter $\theta^\star$, we can easily see that

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

The difference between $\ell(\theta)$ and $\tilde{\ell}_t(\theta)$ does not depend on θ, hence also

$$\bar{\theta}_t \in \operatorname*{argmax}_{\theta \in \Theta} \tilde{\ell}_t(\theta). \tag{19}$$

For $\theta \in \Theta$, define

$$M_t(\theta) := \int_0^t \frac{(b(\theta, X_s) - b(\theta^\star, X_s))}{\sigma(X_s)} dW_s.$$

The process $(M_t(\theta), t \geq 0)$ is a continuous local martingale, with quadratic variation given by

$$A_t(\theta) := < M(\theta) >_t = \int_0^t \frac{(b(\theta, X_s) - b(\theta^\star, X_s))^2}{\sigma^2(X_s)} ds.$$

Note that

$$\tilde{\ell}_t(\theta) = -\frac{1}{2} A_t(\theta) + M_t(\theta), \tag{20}$$

Recall that π, given by the Lemma 1 is an invariant distribution of X and denote

$$\tilde{\ell}(\theta) = -\frac{1}{2}\pi\left(\frac{(b(\theta, .) - b(\theta^\star, .))^2}{\sigma^2(.)}\right) \tag{21}$$

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

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

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

$$\{\frac{1}{t} A_t(\theta)\}_{t>0} \text{ is equicontinuous } \; P - a.s. \tag{22}$$

Indeed,

$$\frac{1}{t}|A_t(\theta) - A_t(\theta')| \leq C|\theta - \theta'|^\kappa \frac{1}{t}\int_0^t K^2(X_s) ds,$$

where $C = 2[Diam(\Theta)]^\kappa$ and K given by the Assumptions 7(i) is sub-polynomial. Using ergodic theorem, which holds thanks to the Lemma1, $\frac{1}{t}\int_0^t K^2(X_s)ds$ converges almost surely to some finite limit. Hence (22) follows. As a consequence,

$$\lim_{t\to\infty} \sup_{\theta\in\Theta} \left| -\frac{1}{2t} A_t(\theta) - \tilde{\ell}(\theta) \right| = 0 \qquad P - a.s. \tag{23}$$

Denote

$$A_t(\theta, \theta') := < M_t(\theta) - M_t(\theta') >_t .$$

Using Assumptions 5 and 7(i), for all $(\theta, \theta') \in \Theta^2$,

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

where $V_t := \int_0^t (\frac{K^2(X_s)}{\sigma^2(X_s)} \vee 1) ds \to \infty, \quad \text{if} \quad t \to \infty$. Therefore all assumptions of the Theorem 2 in Loukianova and Loukianov [2005] are satisfied. As a conclusion, the family $\{\frac{M_t(\theta)}{A_t(\theta)}; \theta \in \Theta, t \geq 0\}$ satisfies the Uniform Law of Large Numbers on any compact $K \in \Theta$ not containing $\theta^\star$, i.e.

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

We deduce, using (23), that

$$\lim_{t\to\infty} \sup_{\theta\in K} \left| \frac{M_t(\theta)}{t} \right| = 0$$

and hence, $P - a.s.$

$$\sup_{\theta\in K} |t^{-1}\tilde{\ell}_t(\theta) - \tilde{\ell}(\theta)| \to 0. \tag{24}$$

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 Van der Vaart [1998] 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

$$\tilde\ell(\theta) \le 0, \quad \tilde\ell(\theta) = 0 \quad \Longleftrightarrow \quad \theta = \theta^\star \tag{25}$$

and hence

$$\sup_{\theta:\ d(\theta,\theta^\star)\ge\varepsilon} \tilde\ell(\theta) < \tilde\ell(\theta^\star) \tag{26}$$

is trivially satisfied in our case. We deduce from (24) and (26) that P-a.s. for all $\varepsilon > 0$,

$$\lim_{t\to\infty} \sup_{d(\theta,\theta^\star)\ge\varepsilon} \frac{1}{t}\tilde\ell_t(\theta) < \tilde\ell(\theta^\star)$$

and hence for $t > t(\omega)$ large enough

$$\sup_{d(\theta,\theta^\star)\ge\varepsilon} \tilde\ell_t(\theta) < \tilde\ell_t(\theta^\star)$$

and finally for $t > t(\omega)$,

$$d(\bar\theta_t, \theta^\star) < \varepsilon,$$

which means the a.s. consistency. □

Recall that I is the Fisher information given by (3).

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 13. *Suppose that Assumptions 1–9 hold. Then the MLE $\bar\theta_t$ is asymptotically normal:*

$$t^{1/2}(\bar\theta_t - \theta^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(\theta^\star)) \quad \textit{as} \quad t\to\infty.$$

Proof. Due to Assumptions 5 and 7, Theorem 2.2 in Hutton and Nelson [1984] and Theorem 1 in Loukianova and Loukianov [2005] for all $t > 0$ the criterion function $\tilde\ell_t(\theta, X)$ is twice continuously differentiable in θ.

From (18) 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

$$\partial_{\theta_i}\tilde\ell_t(\theta) = -\int_0^t \frac{(b(\theta,X_s) - b(\theta^\star,X_s))\partial_{\theta_i} b(\theta,X)}{\sigma^2(X_s)} ds + \int_0^t \frac{\partial_{\theta_i} b(\theta,X_s)}{\sigma(X_s)} dW_s, \tag{27}$$

for $i = 1, \dots, d$. A Taylor expansion around $\bar\theta_t$ yields

$$\int_0^1 \frac{1}{t}\partial^2_\theta \tilde\ell_t(\theta^\star + s(\bar\theta_t - \theta^\star)) ds \times \sqrt{t}(\bar\theta_t - \theta^\star) = -\frac{1}{\sqrt{t}}\nabla_\theta\tilde\ell_t(\theta^\star). \tag{28}$$

Hence, to obtain a CLT for the estimation error $t^{1/2}(\bar\theta_t - \theta^\star)$ we will first show the convergence of the right hand side in (28). The equation (27) gives for $\theta = \theta^\star$

$$\nabla_\theta\tilde\ell_t(\theta^\star) = \int_0^t \frac{\nabla_\theta b(\theta^\star, X_s)}{\sigma(X_s)} dW_s$$

such that the central limit theorem for multidimensional local martingales Küchler and Sørensen [1999] gives

$$t^{-1/2}\nabla_\theta\tilde\ell_t(\theta^\star) = t^{-1/2}\int_0^t \frac{\nabla_\theta b(\theta^\star, X)}{\sigma(X_s)} dW_s \xrightarrow{\mathcal{L}} \mathcal{N}(0, I). \tag{29}$$

In the next step we prove the convergence of

$$\int_0^1 \frac{1}{t}\partial_\theta^2\tilde{\ell}_t(\theta^\star + s(\bar{\theta}_t - \theta^\star))ds.$$

From (27) we see that for $(i,j) \in \{1,\ldots,d\}$,

$$\begin{aligned}\partial^2_{\theta_i\theta_j}\tilde{\ell}_t(\theta) &= -\int_0^t \frac{(b(\theta,X_s)-b(\theta^\star,X_s))\partial^2_{\theta_i,\theta_j}b(\theta,X_s)}{\sigma^2(X_s)}ds - \int_0^t \frac{\partial_{\theta_i}b(\theta,X_s)\partial_{\theta_j}b(\theta,X_s)}{\sigma^2(X_s)}ds \\ &\quad + \int_0^t \frac{\partial^2_{\theta_i\theta_j}b(\theta,X_s)}{\sigma(X_s)}dW_s \\ &:= U_t^1(\theta) + U_t^2(\theta) + U_t^3(\theta). \end{aligned} \tag{30}$$

Using the ergodic theorem, P-a.s.

$$\frac{1}{t}U_t^1(\theta) \to U_\infty^1(\theta) := -\int_{\mathbb{R}} \frac{(b(\theta,x)-b(\theta^\star,x))\partial^2_{\theta_i,\theta_j}b(\theta,x)}{\sigma^2(x)}\pi(dx);$$

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

Moreover, using Assumption 7 and 8 and the same argument which were used to prove the equicontinuity (22) 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 Loukianova and Loukianov [2005] together with Assumptions 5 ,7 and 8 gives that P-a.s.

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

Using (30) and the four last displays we obtain P-a.s.

$$\sup_{\theta\in\Theta}\left|t^{-1}\partial_\theta^2\tilde{\ell}_t(\theta) - (U_\infty^1(\theta) - I(\theta))\right| \to 0 \tag{31}$$

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

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

and

$$t^{-1}\int_0^1 \partial_\theta^2\tilde{\ell}_t(\theta^\star + s(\bar{\theta}_t-\theta^\star))ds \to -I(\theta^\star). \tag{32}$$

Finally, from the non-degeneracy of the Fisher information matrix $I(\theta^\star)$, (29), (32), and Slutsky's theorem, we deduce the asymptotic normality of the estimator. □

6.2. **Local asymptotic normality and efficiency.** To obtain an asymptotic efficiency result in the sense of Hàjek-Le Cam's convolution theorem we prove now the local asymptotic normality property for the statistical experiment $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathcal{P})$. From this result we can then deduce later on efficiency of the discretized estimator with jump filter (cf. Theorem 3 and 4).

Theorem 14. *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,*

$$\ell_t(\theta^\star + \frac{h}{\sqrt{t}}) - \ell_t(\theta^\star) \xrightarrow{\mathcal{L}} -1/2 h^\top I(\theta^\star)h + N, \quad \text{as} \quad t \to \infty, \tag{33}$$

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

Proof.

$$\begin{aligned}
\ell_t(\theta^\star + \frac{h}{\sqrt{t}}) - \ell_t(\theta^\star) &= -\frac{1}{2}\int_0^t \frac{(b(\theta^\star + \frac{h}{\sqrt{t}}, X_s) - b(\theta^\star, X_s))^2 ds}{\sigma^2(X_s)} \\
&+ \int_0^t \frac{(b(\theta^\star + \frac{h}{\sqrt{t}}, X_s) - b(\theta^\star, X_s))}{\sigma(X_s)} dW_s \\
&= -\frac{1}{2}\int_0^1\int_0^1 \left(\frac{1}{t}\int_0^t \frac{h^\top(\nabla b(\theta^\star + \frac{hu}{\sqrt{t}}, X_s)\nabla b^\top(\theta^\star + \frac{hu'}{\sqrt{t}}, X_s)h}{\sigma^2(X_s)} ds\right) du du' \\
&+ \frac{1}{\sqrt{t}}\int_0^t \frac{\nabla b^T(\theta^\star, X_s)h}{\sigma(X_s)} dW_s + R_t.
\end{aligned}$$

Where

$$R_t := \int_0^t \frac{(b(\theta^\star + \frac{h}{\sqrt{t}}, X_s) - b(\theta^\star, X_s))}{\sigma(X_s)} dW_s - \frac{1}{\sqrt{t}}\int_0^t \frac{\nabla b^T(\theta^\star, X_s)h}{\sigma(X_s)} dW_s.$$

Using Assumption 7 and the ergodic theorem, for all fixed $r > 0, r' > 0$ such that $\theta^\star + r \in \Theta$, $\theta^\star + r' \in \Theta$ we obtain

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \frac{h^\top \nabla b(\theta^\star + r, X_s)\nabla b^\top(\theta^\star + r', X_s)h}{\sigma^2(X_s)} ds = \int_{\mathbb{R}} \frac{h^\top \nabla b(\theta^\star + r, x)\nabla b^\top(\theta^\star + r', x)h}{\sigma^2(X_s)} d\pi(x)$$

P-a.s. and Assumption (8) and Lemma 1 imply that this last limit is finite. Moreover, using Assumption 7 it can be shown that this convergence is uniform, hence for $hu/\sqrt{t} \to 0$ it gives that $P - a.s.$

$$\begin{aligned}
\lim_{t\to\infty} \int_0^1 du \int_0^1 du' \frac{1}{t}\int_0^t & \frac{h^\top \nabla b(\theta^\star + \frac{hu}{\sqrt{t}}, X_s)\nabla b^\top(\theta^\star + \frac{hu'}{\sqrt{t}}, X_s)h}{\sigma^2(X_s)} ds \\
&= \int_{\mathbb{R}} \frac{h^\top \nabla b(\theta^\star, x)\nabla b^\top(\theta^\star, x)h}{\sigma^2(X_s)} d\pi(x) = h^\top I(\theta^\star)h.
\end{aligned} \tag{34}$$

Using Markov inequality

$$P(|R_t| \geq \varepsilon) \leq \frac{Var R_t}{\varepsilon^2} \leq \frac{\|h\|^2}{\varepsilon^2}\frac{1}{t}\int_0^t \left(\frac{\|h\|}{\sqrt{t}}\right)^{2\kappa} E\left(\frac{K_1^2(X_s)}{\sigma^2(X_s)}\right) ds, \tag{35}$$

where K_1 is a Holder constant of ∇b is supposed to be at most of polynomial growth. Using ergodic theorem in mean, we obtain $R_t \to 0$ in P probability.

Due to the CLT for martingales in Küchler and Sørensen [1999]

$$\frac{1}{\sqrt{t}}\int_0^t \frac{\nabla b^\top(\theta^\star, X_s)h}{\sigma(X_s)} dW_s \to \mathcal{N}(0, h^\top I(\theta^\star)h)$$

in distribution. Combining the latter equation with (34)–(35), we obtain (33). This implies together with Theorem 13 that $\bar{\theta}_t$ is asymptotically efficient in the sense of the Hájek-Le Cam convolution theorem. □

6.3. **Proofs of Theorems 2, 3 and 4.**

Proof of Theorem 2. Let $\tilde{\ell} : \Theta \to \mathbb{R}$ be given by (21) and define

$$\ell(\theta) = \tilde{\ell}(\theta) + \frac{1}{2}\pi\left(\frac{b^2(\theta^\star, x)}{\sigma^2(x)}\right). \tag{36}$$

Under Assumptions 1 and 5 the last term in the right hand side of (36) is finite.We will apply Wald's method for proving consistency of M estimators (see for example Theorem 5.7 in Van der Vaart [1998]). It follows from (25) that

$$\sup_{\theta;\ d(\theta,\theta^\star)\geq\varepsilon} \ell(\theta) \leq \ell(\theta^\star). \tag{37}$$

Therefore, it remains to prove that

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} |t_n^{-1}\ell_{t_n}^n(\theta) - \ell(\theta)| = 0 \quad \text{in probability.}$$

To obtain this last statement we decompose this difference as follows:

$$\sup_{\theta\in\Theta} |\ell(\theta) - t_n^{-1}\ell_{t_n}^n(\theta)| \leq \sup_{\theta\in\Theta} |\ell(\theta) - t_n^{-1}\ell_{t_n}(\theta)| + \sup_{\theta\in\Theta} |t_n^{-1}(\ell_{t_n}(\theta) - \ell_{t_n}^n(\theta))|. \tag{38}$$

Using respectively the Ergodic Theorem given by Lemma 1 (1) and the Law of Large Numbers for continuous local martingales (Revuz and Yor [1994] p.178) we see that a.s.

$$\frac{1}{t}\int_0^t \frac{b^2(\theta^\star, X_s)}{\sigma^2(X_s)} ds \to \pi\left(\frac{b^2(\theta^\star, x)}{\sigma^2(x)}\right)$$

and

$$\frac{1}{t}\int_0^t \frac{b(\theta^\star, X_s)}{\sigma(X_s)} dW_s \to 0$$

Using these two last display and (24) we see that the first term of the decomposition (38) tends to zero P-a.s.. In order to show the convergence to zero in probability of the second term, we decompose it as follows.

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

Hence, it remains to prove the convergence to zero of $t_n^{-1}|A_n^1(\theta)|$ and $t_n^{-1}|A_n^2(\theta)|$ uniformly in θ. For $t_n^{-1}|A_n^1(\theta)|$ we apply Proposition 6 in the finite activity case and Proposition 7 in the case of infinite activity, together with the fact that $n\Delta_n = O(t_n)$. Indeed, using Assumption 7 and 8 we see that the function $f(\theta,x)=\sigma(x)^{-2}b(\theta,x)^2$ satisfies all assumptions of Propositions 6 or 7. For the second term $t_n^{-1}|A_n^2(\theta)|$ we use Lemma 11. □

Proof of Theorem 3. A Taylor expansion around $\hat\theta_n$ yields

$$\frac{1}{t_n}\int_0^1 \partial_\theta^2 \ell_{t_n}^n(\theta^\star+s(\hat\theta_n-\theta^\star))ds\times t_n^{1/2}(\hat\theta_n-\theta^\star)=-\frac{1}{t_n^{1/2}}\nabla_\theta \ell_{t_n}^n(\theta^\star). \tag{39}$$

For the right hand side we find that

$$\frac{1}{{t_n}^{1/2}}\nabla_\theta \ell_{t_n}^n(\theta^\star)=\frac{\nabla_\theta \ell_{t_n}^n(\theta^\star)-\nabla_\theta \ell_{t_n}(\theta^\star)}{\sqrt{t_n}}+\frac{\nabla_\theta \ell_{t_n}(\theta^\star)}{\sqrt{t_n}}. \tag{40}$$

By (29) we have that under P

$$\frac{\nabla_\theta \ell_{t_n}(\theta^\star)}{\sqrt{t_n}}\xrightarrow{\mathcal{L}} N(0,I(\theta^\star)),\quad n\to\infty. \tag{41}$$

The first term of the sum on the right hand side of (40) has the form

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

By applying Proposition 6 for $k=1,\dots d$ with $f_k(\theta^\star,x)=\sigma(x)^{-2}\partial_{\theta_k}b(\theta^\star,x)$, and using Assumptions 7– 8 we obtain that

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

as $n\to\infty$. Furthermore, Lemma 11 (ii) leads to

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

as $n\to\infty$. Combining now the last three displays results in

$$\frac{\nabla_\theta \ell_{t_n}^n(\theta^\star)-\nabla_\theta \ell_{t_n}(\theta^\star)}{t_n^{1/2}}\xrightarrow{P}0$$

such that (40) and (41) give

$$t_n^{1/2}\nabla_\theta \ell_{t_n}^n(\theta^\star)\xrightarrow{d} N(0,I(\theta^\star)),\quad n\to\infty.$$

To finish the proof it remains to show the convergence of the left hand side in (39). For $(j,k) \in \{1,\dots,d\}^2$ and $\theta \in \Theta$,

$$
\begin{aligned}
& t_n^{-1} \sup_{\theta\in\Theta} \left| \left(\partial^2_{\theta_j\theta_k} \ell^n_{t_n}(\theta) - \partial^2_{\theta_j\theta_k} \ell_{t_n}(\theta) \right) \right| \\
& \leq t_n^{-1} \sup_{\theta\in\Theta} \left| \int_0^{t_n} \sigma(X_s)^{-2} \partial^2_{\theta_j\theta_k} b(\theta, X_s)\, dX^c_s - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \partial^2_{\theta_j\theta_k} b(\theta, X_{t_{i-1}}) \Delta^n_i X \mathbf{1}_{|\Delta^n_i X| \leq v_n} \right| \\
& + t_n^{-1} \sup_{\theta\in\Theta} \left| \int_0^{t_n} \sigma(X_s)^{-2} \partial_{\theta_j} b(\theta, X_s) \partial_{\theta_k} b(\theta, X_s)\, ds - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \partial_{\theta_j} b(\theta, X_s) \partial_{\theta_k} b(\theta, X_s) \Delta^n_i Id \right| \\
& + t_n^{-1} \sup_{\theta\in\Theta} \left| \int_0^{t_n} \sigma(X_s)^{-2} \partial^2_{\theta_j\theta_k} b(\theta, X_s) b(\theta, X_s)\, ds - \sum_{i=1}^n \sigma(X_{t_{i-1}})^{-2} \partial^2_{\theta_j\theta_k} b(\theta, X_s) b(\theta, X_s) \Delta^n_i Id \right| \\
& = U^1_n + U^2_n + U^3_n.
\end{aligned}
$$

Proposition 6 together with Assumptions 7– 8 state that

$$
U^1_n \xrightarrow{P} 0, \quad \text{as } n \to \infty. \tag{42}
$$

Lemma 11 (i) gives for k=2,3

$$
U^k_n \xrightarrow{P} 0, \quad \text{as } n \to \infty. \tag{43}
$$

Combining (42) and (43) with consistency of $\hat\theta$ and $\bar\theta$ we get

$$
\int_0^1 \frac{1}{t_n} |\partial^2_\theta \ell^n_{t_n}(\theta^\star + s(\hat\theta_n - \theta^\star)) - \partial^2_\theta \ell_{t_n}(\theta^\star + s(\bar\theta_{t_n} - \theta^\star))| ds \xrightarrow{P} 0,
$$

and hence, using (32)

$$
\frac{1}{t_n} \int_0^1 \partial^2_\theta \ell^n_{t_n}(\theta^\star + s(\hat\theta_n - \theta^\star)) ds \xrightarrow{P} -I(\theta^\star)
$$

as $n \to \infty$ such that the result follows. □

Proof of Theorem 4. By replacing in the previous proof Proposition 6 by Proposition 7 we obtain the result for the infinite activity case. □

7. Proofs for jump filtering

In this section we prove the results that were used in the Section 4 to obtain the convergence of the jump filter (cf. Proposition 6 and 7) to integral functionals with respect to the continuous martingale part of X. We start by proving the Lemma 8 that shows the convergence of the jump filter approximation to the continuous part in the finite activity case.

We recall some notations: μ 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 such that we have $L_t = \int_0^t \int_{\mathbb{R}} z\mu(ds, dz)$. In the specific situation where the Lévy process L has a finite intensity $\nu(\mathbb{R}) < \infty$, we shall denote by $N_t = \int_0^t \int_{\mathbb{R}} \mu(ds, dz)$ the process that counts the number of jumps up to time t.

Proof of Lemma 8. For all $n \in \mathbb{N}^*$, $i \in \mathbb{N}^*$ we define the set where increments of X are small:

$$
K^i_n = \{|\Delta^n_i X| \leq v_n\}, \tag{44}
$$

the event that L and so also X do not jump:

$$M_n^i = \{\Delta_i^n N = 0\}, \tag{45}$$

and the event that an increment of the jump part is small:

$$D_n^i = \left\{ \left| \Delta_i^n X^J \right| \leq \frac{v_n}{3} \right\}, \tag{46}$$

where we denoted by X^J the jump part of X given by

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

We start by proving (i). Using the previously defined sets we introduce the following quantities.

$$G_n^1(\theta) := \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^J\right) 1_{K_n^i \cap (M_n^i)^c}, \tag{47}$$

$$G_n^2(\theta) := \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c\right) 1_{(K_n^i)^c \cap D_n^i}, \tag{48}$$

$$G_n^3(\theta) := \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c\right) 1_{(K_n^i)^c \cap (D_n^i)^c}, \tag{49}$$

and decompose the difference to be estimated as follows:

$$\sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n}\right) = G_n^1(\theta) + G_n^2(\theta) + G_n^3(\theta). \tag{50}$$

To prove the convergence of $G_n^1(\theta)$ we decompose the set $K_n^i \cap (M_n^i)^c$ into three disjoint events

$$\begin{aligned} \mathbf{1}_{K_n^i \cap (M_n^i)^c} &= \mathbf{1}_{\{\Delta_i^n N \geq 2\} \cap K_n^i} + \mathbf{1}_{\{\Delta_i^n N = 1, |\Delta_i^n L| \geq 2v_n/\gamma_{min}\} \cap K_n^i} \\ &\quad + \mathbf{1}_{\{\Delta_i^n N = 1, |\Delta_i^n L| < 2v_n/\gamma_{min}\} \cap K_n^i} \end{aligned} \tag{51}$$

Using Lemma 15 (3), the definition of v_n and Markov's inequality we can see that the second indicator of this decomposition is on an event that has small probability. Indeed, for all $p > 1$,

$$P\left(\{\Delta_i^n N = 1, |\Delta_i^n L| \geq 2v_n/\gamma_{min}\} \cap K_n^i\right) \leq P\left(|\Delta_i^n X^c| \geq v_n\right) = O(\Delta_n^{p/2} v_n^{-p}) = O(\Delta_n^{\varepsilon p}).$$

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

$$\begin{aligned} E\left|\Delta_i^n X^J\right|^2 &\leq 2E\left|\int_{t_{i-1}}^{t_i} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s-}) z \tilde{\mu}(ds, dz)\right|^2 + 2E\left[\int_{t_{i-1}}^{t_i} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_s) z ds \nu(dz)\right]^2 \\ &\leq 2\int_{t_{i-1}}^{t_i} \int_{\mathbb{R}\setminus\{0\}} E[\gamma^2(X_s)] z^2 ds \nu(dz) + 2\int_{t_{i-1}}^{t_i} \int_{\mathbb{R}\setminus\{0\}} E[\gamma^2(X_s)] |z| ds \nu(dz) \int_{t_{i-1}}^{t_i} \int_{\mathbb{R}\setminus\{0\}} |z| ds \nu(dz) \\ &= O(\Delta_n), \end{aligned} \tag{52}$$

where in the last line we have used Assumption 1, Assumption 3 (i), Assumption 4 (ii) and Lemma 1 statement (3). Using Hölder's inequality twice and Lemma 1 statement (3) we get for all $p > 0$,

$$E \sup_{\theta\in\Theta} \left| \sum_{i=1}^{n} f(\theta, X_{t_{i-1}}) \Delta_i^n X^J 1_{\{\Delta_i^n N=1, |\Delta_i^n L| \geq 2v_n/\gamma_{\min}\} \cap K_n^i} \right| = O(n\Delta_n^{\varepsilon p}). \tag{53}$$

For the third indicator function in (51) we observe that

$$\begin{aligned} &E \sup_{\theta\in\Theta} |\sum_{i=1}^{n} f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^J\right) 1_{\{\Delta_i^n N=1, |\Delta_i^n L| < 2v_n/\gamma_{\min}\} \cap K_n^i} \\ &\leq \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_{|z|<2v_n/\gamma_{\min}} E\left[\sup_{\theta\in\Theta} |f(\theta, X_{t_{i-1}})\gamma(X_s)|\right] |z| ds\nu(dz) \\ &= O(n\Delta_n \int_{|z|<2v_n/\gamma_{\min}} |z|\nu(dz)), \end{aligned} \tag{54}$$

where we have used the sub-polynomial growth of γ, f and Lemma 1 statement(3).

For the first indicator in (51) we obtain by Hölder's inequality with conjugated exponents, p, q, such that $p^{-1} + q^{-1} = 1$, and $q^{-1} = 1 - \varepsilon/2$,

$$\begin{aligned} &E \sup_{\theta\in\Theta} \left| \sum_{i=1}^{n} f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n}\right) 1_{\{\Delta_i^n N \geq 2\} \cap K_n^i} \right| \\ &\leq \sum_{i=1}^{n} E \sup_{\theta\in\Theta} |f(\theta, X_{t_{i-1}})| \left(|\Delta_i^n X^c| + v_n\right) 1_{\{\Delta_i^n N \geq 2\}} \\ &\leq \sum_{i=1}^{n} \left(E \sup_{\theta\in\Theta} |f(\theta, X_{t_{i-1}})|^p (|\Delta_i^n X^c| + v_n)^p \right)^{1/p} P(\Delta_i^n N \geq 2)^{1/q} \\ &= O(n(\Delta_n^{1/2} + v_n)\Delta_n^{2/q} = O(n\Delta_n^{5/2-2\varepsilon}), \end{aligned} \tag{55}$$

where we have used that $P(\Delta_i^n N \geq 2) = O(\Delta_n^2)$.

From (53), (54) and (55) it follows that

$$E \sup_{\theta\in\Theta} |G_n^1(\theta)| \leq O(n\Delta_n \int_{|z|\leq 2v_n/\gamma_{min}} |z|\nu(dz)) + O(n\Delta_n^{5/2-2\varepsilon}). \tag{56}$$

To estimate $G_n^2(\theta)$ note first that for any $p > 1$,

$$P\left((K_n^i)^c \cap D_n^i\right) \leq P\left(|\Delta_i^n X^c| > 2v_n/3\right) = O(\Delta_n^{\varepsilon p}).$$

Hence, by using Hölder's inequality, sub-polynomial growth of f, (3) of Lemma 1 and (3) of Lemma 15 we obtain for any $p > 1$,

$$E \sup_{\theta\in\Theta} |G_n^2(\theta)| = E \sup_{\theta\in\Theta} |\sum_{i=1}^{n} f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c\right) 1_{(K_n^i)^c \cap D_n^i}| = O(n\Delta_n^{\varepsilon p}). \tag{57}$$

To estimate $G_n^3(\theta)$ note first that

$$\begin{aligned}
&P((D_n^i)^c) = P(|\Delta_i^n X^J| > v_n/3) \\
&\leq P(|\int_{t_{i-1}}^{t_i}\int_{|z|\geq v_n}\gamma(X_{s-})z\mu(ds,dz)| > v_n/6) + P(|\int_{t_{i-1}}^{t_i}\int_{|z|<v_n}\gamma(X_{s-})z\mu(ds,dz)| > v_n/6) \\
&\leq P(|\int_{t_{i-1}}^{t_i}\int_{|z|\geq v_n}\gamma(X_{s-})z\mu(ds,dz)| > 0) + \frac{6}{v_n}E[|\int_{t_{i-1}}^{t_i}\int_{|z|<v_n}\gamma(X_s)zds\nu(dz)|] \\
&\leq P(\int_{t_{i-1}}^{t_i}\int_{|z|\geq v_n}\mu(ds,dz) \geq 1) + \frac{6}{v_n}E[|\int_{t_{i-1}}^{t_i}\int_{|z|<v_n}\gamma(X_s)zds\nu(dz)|] \\
&= 1 - \exp\left(-\int_{t_{i-1}}^{t_i}\int_{|z|\geq v_n}ds\nu(dz)\right) + \frac{6}{v_n}E[|\int_{t_{i-1}}^{t_i}\int_{|z|<v_n}\gamma(X_s)zds\nu(dz)|] \\
&= O\left(\Delta_n(\int_{|z|\geq v_n}\nu(dz) + \frac{1}{v_n}\int_{|z|<v_n}|z|\nu(dz))\right).
\end{aligned}$$

Hence, using Hölder's inequality, the assumptions on f and (3) of Lemma 15 we obtain for any $q > 1$ that

$$\begin{aligned}
E\sup_{\theta\in\Theta}|G_n^3(\theta)| &\leq E\sum_{i=1}^n \sup_{\theta\in\Theta}|f(\theta, X_{t_{i-1}})||\Delta_i^n X^c|1_{\{|\Delta_i^n X|>v_n,(D_n^i)^c\}} \leq O(n\Delta_n^{1/2})P((D_n^i)^c)^{1/q} \\
&\leq O(n\Delta_n^{1/2})\Delta_n^{1/q}\left(\int_{|z|\geq v_n}\nu(dz) + \frac{1}{v_n}\int_{|z|<v_n}|z|\nu(dz)\right)^{1/q}. \qquad (58)
\end{aligned}$$

Finally, choosing $q^{-1} = 1 - \varepsilon/2$, we get from (56), (57) and (58) that

$$\begin{aligned}
&E\left[\sup_{\theta\in\Theta}\left|\sum_{i=1}^n f(\theta, X_{t_{i-1}})\left(\Delta_i^n X^c - \Delta_i^n X\mathbf{1}_{|\Delta_i^n X|\leq v_n}\right)\right|\right] \leq O(n\Delta_n\int_{|z|\leq 2v_n/\gamma_{min}}|z|\nu(dz)) \\
&+ O(n\Delta_n^{5/2-2\varepsilon}) + O(n\Delta_n^{3/2-\varepsilon/2})(\int_{|z|\geq v_n}\nu(dz) + \frac{1}{v_n}\int_{|z|<v_n}|z|\nu(dz))^{1-\varepsilon/2}
\end{aligned}$$

In particular, using the definition of v_n, finiteness of ν and of its first moment we immediately get

$$E\sup_{\theta\in\Theta}\left|\sum_{i=1}^n f(\theta, X_{t_{i-1}})\left(\Delta_i^n X^c - \Delta_i^n X\mathbf{1}_{|\Delta_i^n X|\leq v_n}\right)\right| = O(n\Delta_n^{3/2-\varepsilon/2}),$$

hence (i) is proved.

To prove (ii) we decompose the approximation by the jump filter as follows:

$$\sum_{i=1}^n f(\theta, X_{t_{i-1}})\left(\Delta_i^n X^c - \Delta_i^n X\mathbf{1}_{|\Delta_i^n X|\leq v_n}\right) = G_n^1(\theta) + A_n^2(\theta) + A_n^3(\theta), \qquad (59)$$

where $G_n^1(\theta)$ is given by (47) and

$$A_n^2(\theta) := \sum_{i=1}^n f(\theta, X_{t_{i-1}})\,(\Delta_i^n X^c)\,1_{(K_n^i)^c \cap (M_n^i)^c} \tag{60}$$

$$A_n^3(\theta) := \sum_{i=1}^n f(\theta, X_{t_{i-1}})\,(\Delta_i^n X^c)\,1_{(K_n^i)^c \cap M_n^i}.$$

Observe that

$$A_n^2(\theta) = \sum_{i=1}^n f(\theta, X_{t_{i-1}})\Delta_i^n X^c \mathbf{1}_{(M_n^i)^c} - \sum_{i=1}^n f(\theta, X_{t_{i-1}})\Delta_i^n X^c \mathbf{1}_{\{|\Delta_i^n X| \le v_n\} \cap (M_n^i)^c}. \tag{61}$$

We first show that after suitable renormalization the first term of this decomposition converges to zero in probability. Let $e_i := f(\theta, X_{t_{i-1}})\Delta_i^n X^c \mathbf{1}_{(M_n^i)^c}$. Denote $\mathcal{F}_i = \sigma\{(W_s)_{0<s\le t_i}, (L_s)_{0<s\le t_i}, X_0\}$, then

$$E[e_i|\mathcal{F}_{i-1}] = f(\theta, X_{t_{i-1}})E\left[\int_{t_{i-1}}^{t_i} \sigma(X_s)dW_s 1_{(M_n^i)^c}|\mathcal{F}_{i-1}\right]$$
$$+ f(\theta, X_{t_{i-1}})E\left[\int_{t_{i-1}}^{t_i} b(\theta^\star, X_s)ds 1_{(M_n^i)^c}|\mathcal{F}_{i-1}\right]$$

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

$$E\left[\int_{t_{i-1}}^{t_i} \sigma(X_s)dW_s 1_{(M_n^i)^c}|\mathcal{F}_{i-1}\right] = E\left[1_{(M_n^i)^c} E\left[\int_{t_{i-1}}^{t_i} \sigma(X_s)dW_s|\mathcal{F}_{i-1} \vee \sigma(L)\right]|\mathcal{F}_{i-1}\right] = 0$$

and so

$$|E[e_i|\mathcal{F}_{i-1}]| \le |f(\theta, X_{t_{i-1}})| \int_{t_{i-1}}^{t_i} E\left[|b(\theta^\star, X_s)|\,1_{(M_n^i)^c}|\mathcal{F}_{i-1}\right] ds. \tag{62}$$

Recall that

$$P\left((M_n^i)^c\right) = 1 - P(\Delta_i^n N = 0) = O(\Delta_n).$$

Using Hölder inequality, Lipshitz continuity of $b(\theta^\star, .)$, the continuity of its Lipshitz constant given by the Assumption 1 and Lemma 15 (2) we can write for p, q such that $p^{-1} + q^{-1} = 1$, $p \ge 2$ and $C > 0$,

$$E\left[|b(\theta^\star, X_s)|\,1_{(M_n^i)^c}|\mathcal{F}_{i-1}\right] \le \left(E\left[|b(\theta^\star, X_s)|^p\,|\mathcal{F}_{i-1}\right]\right)^{1/p} \Delta_n^{1/q}$$
$$\le C\left(E[|b(\theta^\star, X_s) - b(\theta^\star, X_{t_{i-1}})|^p|\mathcal{F}_{t_{i-1}}] + |b(\theta^\star, X_{t_{i-1}})|^p\right)^{1/p} \Delta_n^{1/q}$$
$$\le C\left(\left(E\left[|X_s - X_{t_{i-1}}|^p|\mathcal{F}_{t_{i-1}}\right]\right)^{1/p} + |b(\theta^\star, X_{t_{i-1}})|\right)\Delta_n^{1/q}$$
$$\le C\Delta_n^{1/q}\left(\Delta_n^{1/p}(1 + |X_{t_{i-1}}|^p)^{1/p} + |b(\theta^\star, X_{t_{i-1}})|\right). \tag{63}$$

Using the fact that $b(\theta^\star, .)$ and $\sup_{\theta\in\Theta} |f(\theta, .)|$ are sub-polynomial and choosing again $1/q = 1 - \varepsilon/2$, (which also guarantees $p > 2$,) we obtain

$$|E[e_i|\mathcal{F}_{i-1}]| \le h(|X_{t_{i-1}}|)\Delta_n^{2-\varepsilon/2}, \tag{64}$$

where h is a polynomial function. Finally this implies that under the condition $n\Delta_n^{3-\varepsilon} \to 0$,

$$E\left[\sum_{i=1}^{n}\left|E\left[\frac{e_i}{\sqrt{n\Delta_n}}|\mathcal{F}_{i-1}\right]\right|\right] = O(n^{1/2}\Delta_n^{3/2-\varepsilon/2}) \to 0. \tag{65}$$

Next, we bound the moment of order two of e_i.

By Hölder's inequality with $1/q = 1-\varepsilon/2$, $1/p = 1-1/q$, we have

$$\begin{aligned} E[e_i^2] &\leq E\left[f(\theta, X_{t_{i-1}})^{2p}(\Delta_i^n X^c)^{2p}\right]^{1/p} P\left[(M_n^i)^c\right]^{1/q} \\ &\leq \Delta_n P\left[(M_n^i)^c\right]^{1-\varepsilon/2} = O(\Delta_n^{2-\varepsilon/2}), \end{aligned} \tag{66}$$

where in the last line we used again Hölder's inequality, the sub-linear growth of f, together with Lemma 15 (3). Hence,

$$E\left[\left|\sum_{i=1}^{n} E\left[\left(\frac{e_i}{\sqrt{n\Delta_n}}\right)^2|\mathcal{F}_i\right]\right|\right] = \sum_{i=1}^{n} E\left[\left(\frac{e_i}{\sqrt{n\Delta_n}}\right)^2\right] = O(\Delta_n^{1-\varepsilon/2}) \to 0. \tag{67}$$

Under (65) and (67) we obtain from Lemma 9 in Genon-Catalot and Jacod [1993] that

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n} f(\theta, X_{t_{i-1}})\Delta_i^n X^c \mathbf{1}_{(M_n^i)^c} = \sum_{i=1}^{n}\frac{e_i}{\sqrt{n\Delta_n}} \xrightarrow{P} 0 \tag{68}$$

if $n\Delta_n^{3-\varepsilon} \to 0$. Recall that the second term in the decomposition (61) of A_n^2 is given by

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

We will now bound this term in $\mathbb{L}^1$. We use again the decomposition (51) of $K_n^i \cap (M_n^i)^c$. We find that by computations similar to (55) and (53) respectively, we have

$$E\left|\sum_{i=1}^{n} f(\theta, X_{t_{i-1}})\Delta_i^n X^c \mathbf{1}_{\{\Delta_i^n N \geq 2\}\cap K_n^i}\right| = O(n\Delta_n^{5/2-2\varepsilon}), \tag{69}$$

$$E\left|\sum_{i=1}^{n} f(\theta, X_{t_{i-1}})\Delta_i^n X^c \mathbf{1}_{\{\Delta_i^n N = 1, |\Delta_i^n L| \geq 2v_n/\gamma_{min}\}\cap K_n^i}\right| = O(n\Delta_n^{\varepsilon p}) \tag{70}$$

Moreover, we have that $P(\Delta_i^n N = 1, |\Delta_i^n L| < 2v_n/\gamma_{min}) = P(\int_{t_{i-1}}^{t_i}\int_{|z|<2v_n/\gamma_{min}}\mu(ds, dz) = 1) \leq \Delta_n \int_{|z|<2v_n}\nu(dz)$, where we used $\gamma_{min} \geq 1$. From this, we can easily get

$$E\left|\sum_{i=1}^{n} f(\theta, X_{t_{i-1}})\Delta_i^n X^c \mathbf{1}_{\{\Delta_i^n N=1, |\Delta_i^n L| < 2v_n/\gamma_{min}\}\cap K_n^i}\right| = O\left(n\Delta_n^{3/2-\varepsilon/2}\left(\int_{|z|<2v_n}\nu(dz)\right)^{1-\varepsilon/2}\right). \tag{71}$$

From (61), (68), (69)–(71), we deduce that if $n\Delta_n^{3-\varepsilon} \to 0$,

$$A_n^2(\theta) = o_P(\sqrt{n\Delta_n}) + O_{\mathbb{L}^1}\left(n\Delta_n^{5/2-2\varepsilon} + \left(\int_{|z|<2v_n}\nu(dz)\right)^{1-\varepsilon/2} n\Delta_n^{3/2-\varepsilon/2}\right). \tag{72}$$

It follows immediately from Lemma 15 (3) that for any $p > 1$,

$$P((K_n^i)^c \cap M_n^i) \leq P(|\Delta_i^n X^c| > v_n) = O(\Delta_n^{\varepsilon p}).$$

Hence, using again Hölder's inequality and Lemma 15 (3) again, we see that for any $p>1$,

$$E\left|A_n^3(\theta)\right| = E\left|\sum_{i=1}^n f(\theta,X_{t_{i-1}})\left(\Delta_i^n X^c - \Delta_i^n X\mathbf{1}_{|\Delta_i^n X|\le v_n}\right)1_{(K_n^i)^c\cap M_n^i}\right| = O(n\Delta_n^{\varepsilon p}). \tag{73}$$

Finally, from (56), (72), (73) we obtain that for any $\theta\in\Theta$, if $n\Delta_n^{3-\varepsilon}\to 0$,

$$\sum_{i=1}^n f(\theta,X_{t_{i-1}})\left(\Delta_i^n X^c - \Delta_i^n X\mathbf{1}_{|\Delta_i^n X|\le v_n}\right) =$$

$$o_P(\sqrt{n\Delta_n}) + O_{\mathbb{L}^1}\left(n\Delta_n^{5/2-2\varepsilon} + \left(\int_{|z|<2v_n}\nu(dz)\right)^{1-\varepsilon/2} n\Delta_n^{3/2-\varepsilon/2} + n\Delta_n\int_{|z|\le 2v_n}|z|\nu(dz)\right).$$

This proves (ii). $\square$

Proof of Lemma 9. We start by proving (i). In the infinite jump activity case, the Lévy process has infinite number of jumps on all compact intervals. Hence, it is impossible to introduce the events that the process had no jump, one jump, or more than two jumps on $(t_{i-1},t_i]$ as it was done in the proof of Lemma 8.

Here, we define the event on which all the jumps of L are small :

$$N_n^i = \{|\Delta L_s|\le 3v_n/\gamma_{min};\ \forall s\in(t_{i-1},t_i]\}, \tag{74}$$

where $\Delta L_s := L_s - L_{s-}$. Using the sets K_n^i and D_n^i from (44), we define

$$B_n^1(\theta) := \sum_{i=1}^n f(\theta,X_{t_{i-1}})\left(\Delta_i^n X^J\right)1_{K_n^i\cap(N_n^i)^c};\quad B_n^2(\theta) := \sum_{i=1}^n f(\theta,X_{t_{i-1}})\left(\Delta_i^n X^J\right)1_{K_n^i\cap N_n^i}; \tag{75}$$

and decompose the difference as follows

$$\sum_{i=1}^n f(\theta,X_{t_{i-1}})\left(\Delta_i^n X^c - \Delta_i X\mathbf{1}_{|\Delta_i^n X|\le v_n}\right) = B_n^1(\theta)+B_n^2(\theta)+G_n^2(\theta)+G_n^3(\theta), \tag{76}$$

where $G_n^2(\theta)$ and $G_n^3(\theta)$ are defined in (48)–(49). We start by studying the convergence of $B_n^1(\theta)$. 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

$$A_n^i = \left\{\sum_{t_{i-1}<s\le t_i; s\ne T_i^*}|\Delta L_s| \le \frac{v_n}{\gamma_{max}}\right\}, \tag{77}$$

where γ_{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

$$\begin{aligned}&K_n^i\cap(N_n^i)^c\cap A_n^i\\ &\subset\{|\Delta_i^n X^c+\gamma(X_{T_i^*-})\Delta L_{T_i^*}+\sum_{s\ne T_i^*}\Delta X_s|\le v_n;\quad |\gamma(X_{T_i^*-})\Delta L_{T_i^*}|>3v_n;\quad |\sum_{s\ne T_i^*}\Delta X_s|\le v_n\}\\ &\subset\{|\Delta_i^n X^c|\ge v_n\}.\end{aligned}$$

Hence, using (3) from Lemma 15 we get for all $p>1$:

$$P(K_n^i \cap (N_n^i)^c \cap A_n^i) \leq P\left(|\Delta_i^n X^c| \geq v_n\right) = O(\Delta_n^{p/2} v_n^{-p}) = O(\Delta_n^{\varepsilon p}). \tag{78}$$

Together with (52), which is still true in the infinite activity case, Hölder's inequality, sub-polynomial growth of f and (3) from Lemma 1 this gives for any $p>1$ that

$$E \sup_{\theta\in\Theta} \Big| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^J\right) \left(1_{K_n^i \cap (N_n^i)^c \cap A_n^i}\right) \Big| = O(n\Delta_n^{\varepsilon p}). \tag{79}$$

Using Hölder inequality, sub-polynomial growth of f, Lemma 1 (3), and Lemma 16, we get for $1/p+1/q=1$ and some $C>0$,

$$\begin{aligned}
&E \sup_{\theta\in\Theta} \Big| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c - \Delta_i^n X \mathbf{1}_{|\Delta_i^n X| \leq v_n}\right) 1_{K_n^i \cap (N_n^i)^c \cap (A_n^i)^c} \Big| \\
&\leq \sum_{i=1}^n \left(E \sup_{\theta\in\Theta} |f(\theta, X_{t_{i-1}})|^p (|\Delta_i^n X^c| + v_n)^p \right)^{1/p} \left(P((N_n^i)^c \cap (A_n^i)^c) \right)^{1/q} \\
&\leq Cn(\Delta_n^{1/2} + v_n) \left(\frac{\Delta_n^2}{v_n} \int_{|z|\geq 3v_n/\gamma_{min}} \nu(dz) \right)^{1/q} \\
&\leq Cn v_n^{\varepsilon/2} \Delta_n^{2-\varepsilon} \left(\int_{|z|\geq 3v_n/\gamma_{min}} \nu(dz) \right)^{1-\varepsilon/2}, \quad \text{choosing } 1/q = 1-\varepsilon/2.
\end{aligned} \tag{80}$$

From (79) and (80), we get

$$E \sup_{\theta\in\Theta} |B_n^1(\theta)| = o\left(n\Delta_n^{3/2-\varepsilon/2} \left(\int_{|z|\geq v_n/\gamma_{min}} \nu(dz) \right)^{1-\varepsilon/2} \right). \tag{81}$$

To estimate $B_n^2(\theta)$ we use the bound

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

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

$$E \sup_{\theta\in\Theta} |B_n^2(\theta)| = O(n\Delta_n \int_{|z|\leq 3v_n/\gamma_{min}} |z| \nu(dz)) \leq O(n\Delta_n \int_{|z|\leq 3v_n} |z| \nu(dz)). \tag{82}$$

The $\mathbb{L}^1$ norms of $\sup_{\theta\in\Theta} |G_n^2(\theta)|$ and $\sup_{\theta\in\Theta} |G_n^3(\theta)|$ have been studied in the Lemma 8, when the Lévy process has finite activity. However, the proofs of the upper bounds (57) and (58), obtained in Lemma 8, do not use the fact that $\nu(\mathbb{R}) < \infty$.

Finally, collecting (57), (58) with $1/q = 1-\varepsilon/2$, (81), and (82) we obtain (i). We continue with the proof of (ii). Using the events K_n^i and N_n^i given by (44) and (74) we define

$$B_n^3(\theta) := \sum_{i=1}^n f(\theta, X_{t_{i-1}})\,(\Delta_i^n X^c)\,1_{(K_n^i)^c \cap (N_n^i)^c}; \quad B_n^4(\theta) := \sum_{i=1}^n f(\theta, X_{t_{i-1}})\,(\Delta_i^n X^c)\,1_{(K_n^i)^c \cap (N_n^i)};$$

and decompose the difference as follows

$$\sum_{i=1}^n f(\theta, X_{t_{i-1}})\left(\Delta_i^n X^c - \Delta_i X \mathbf{1}_{|\Delta_i^n X| \le v_n}\right) = B_n^1(\theta) + B_n^2(\theta) + B_n^3(\theta) + B_n^4(\theta), \tag{83}$$

where $B_n^1(\theta)$ and $B_n^2(\theta)$ are given by (75). Using (79) and (80) we can see that

$$E|B_n^1(\theta)| = o\left(n\Delta_n^{2-\varepsilon}\left(\int_{|z|\ge 3v_n/\gamma_{min}} \nu(dz)\right)^{1-\varepsilon/2}\right), \tag{84}$$

while (82) gives the bound for $E|B_n^2(\theta)|$. The role of the event N_n^i (all the jumps of L are small) in the case of the infinite activity is similar to the role of M_n^i (L does not jump) in the finite activity case. Therefore, to estimate $B_n^3(\theta)$ we use a decomposition similar to (61), where we replace M_n^i by N_n^i which leads to

$$B_n^3(\theta) = \sum_{i=1}^n f(\theta, X_{t_{i-1}})\,(\Delta_i^n X^c)\,1_{(N_n^i)^c} - \sum_{i=1}^n f(\theta, X_{t_{i-1}})\,(\Delta_i^n X^c)\,1_{K_n^i \cap (N_n^i)^c}. \tag{85}$$

We will show that the first term of this decomposition goes to zero after suitable normalization. Let $\tilde{e}_i := f(\theta, X_{t_{i-1}})\Delta_i^n X^c \mathbf{1}_{(N_n^i)^c}$. Recall that

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

Therefore, the same arguments that were used to obtain (64) give here

$$|E[\tilde{e}_i|\mathcal{F}_{i-1}]| \le h(|X_{t_{i-1}}|)\Delta_n^{2-\varepsilon/2}\left(\int_{|z|>3v_n/\gamma_{min}} \nu(dz)\right)^{1-\varepsilon/2}, \tag{86}$$

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

$$E\left[\sum_{i=1}^n \left|E\left[\frac{\tilde{e}_i}{\sqrt{n\Delta_n}}|\mathcal{F}_{i-1}\right]\right|\right] = O\left(n^{1/2}\Delta_n^{3/2-\varepsilon/2}\left(\int_{|z|>3v_n/\gamma_{min}} \nu(dz)\right)^{1-\varepsilon/2}\right) \to 0. \tag{87}$$

Next, we bound the second moment of $\tilde{e}_i$. Similarly to (66) we obtain

$$E[\tilde{e}_i^2] \le \Delta_n P[(N_n^i)^c]^{1-\varepsilon/2} = O\left(\Delta_n^{2-\varepsilon/2}\left(\int_{|z|>3v_n/\gamma_{min}} \nu(dz)\right)^{1-\varepsilon/2}\right), \tag{88}$$

Hence, using $\Delta_n \int_{|z|>3v_n/\gamma_{min}} \nu(dz) \to 0$, which is implied by $n\Delta_n^{3-\varepsilon}\left(\int_{|z|>3v_n/\gamma_{min}} \nu(dz)\right)^{2-\varepsilon} \to 0$, we have

$$E\left[\left|\sum_{i=1}^{n} E\left[\left(\frac{\tilde{e}_i}{\sqrt{n\Delta_n}}\right)^2 |\mathcal{F}_i\right]\right|\right] = \sum_{i=1}^{n} E\left[\left(\frac{\tilde{e}_i}{\sqrt{n\Delta_n}}\right)^2\right]$$

$$= O\left(\Delta_n^{1-\varepsilon/2}\left(\int_{|z|>3v_n/\gamma_{min}} \nu(dz)\right)^{1-\varepsilon/2}\right) \to 0. \tag{89}$$

Under (87) and (89) we obtain from Lemma 9 in Genon-Catalot and Jacod [1993] that

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^{n} f(\theta, X_{t_{i-1}})\Delta_i^n X^c \mathbf{1}_{(N_n^i)^c} = \sum_{i=1}^{n} \frac{\tilde{e}_i}{\sqrt{n\Delta_n}} \xrightarrow{P} 0 \tag{90}$$

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

Recall that the second term in the decomposition (85) of B_n^3 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 $\mathbb{L}^1$. Using the set A_n^i defined by (77) we decompose

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

The first term of this decomposition is bounded in $\mathbb{L}^1$ using (78). As a result, for all $p > 1$,

$$E \sup_{\theta \in \Theta} \left| \sum_{i=1}^{n} f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c\right) \left(1_{K_n^i \cap (N_n^i)^c \cap A_n^i}\right) \right| = O(n\Delta_n^{\varepsilon p}). \tag{91}$$

Then, exactly as in (80), we get

$$E\left| \sum_{i=1}^{n} f(\theta, X_{t_{i-1}}) \left(\Delta_i^n X^c\right) \left(1_{K_n^i \cap (N_n^i)^c \cap (A_n^i)^c}\right) \right| = o\left(n\Delta_n^{2-\varepsilon}\left(\int_{|z|\geq v_n/\gamma_{min}} \nu(dz)\right)^{1-\varepsilon/2}\right). \tag{92}$$

As a result,

$$B_n^3(\theta) = o_P(\sqrt{n\Delta_n}) + o_{\mathbb{L}^1}\left(n\Delta_n^{2-\varepsilon}\left(\int_{|z|\geq v_n/\gamma_{min}} \nu(dz)\right)^{1-\varepsilon/2}\right). \tag{93}$$

It remains to estimate the term B_n^4 in the decomposition (83). Observe that for all $p > 1$,

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

where $C > 0$. Using Hölder's inequality twice, this last bound, sub-polynomial growth of f and Lemma 15 (iii) we can easily see that with $1/q = 1 - \varepsilon/2$ we get

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

Finally, collecting (82), (84), (93) and (94) we obtain assertion (ii) of the lemma. □

Proof of Lemma 10. Using $dX_s^c = b(\theta^\star, X_s)ds + \sigma(X_s)dW_s$ we decompose the difference as

$$
\int_0^{t_n} f(\theta, X_s)\; dX_s^c - \sum_{i=1}^n f(\theta, X_{t_{i-1}})\Delta_i^n X^c = A_{n,1}(\theta) + A_{n,2}(\theta) + A_{n,3}(\theta), \tag{95}
$$

where

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

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

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

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 σ and $\sup_\theta |f'(\theta, .)|$, Hölder's inequality, (3) of the Lemma 1 and (1) of the Lemma 15 yields

$$
E\left[(X_s - X_{t_{i-1}})^2 f'^2(\theta, \tilde{x})\sigma^2(X_s)\right] \leq CE[|X_s - X_{t_{i-1}}|^{2q}]^{1/q} \leq C\Delta_n^{1/q}, \tag{98}
$$

where $q > 1$ and C is a positive constant. Hence, for all $\theta \in \Theta$,

$$
E[A^2_{n,1}(\theta)] \leq Cn\Delta_n^{1+1/q}
$$

and consequently

$$\frac{1}{\sqrt{n\Delta_n}} A_{n,1}(\theta) \xrightarrow{L^2} 0. \tag{99}$$

Using Lipshitz continuity of b, and the same arguments than for obtaining (98), it follows immediately that

$$E[\sup_{\theta\in\Theta} |A_{n,2}(\theta)|] \leq Cn\Delta_n^{1+1/q} \tag{100}$$

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

$$\frac{1}{\sqrt{n\Delta_n}} \sup_{\theta\in\Theta} |A_{n,2}(\theta)| \xrightarrow{L^1} 0. \tag{101}$$

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

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

Denote

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

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

$$E\left[e_i^n | \mathcal{F}_{t_{i-1}}\right] = 0.$$

Using Hölder's inequality and isometry property of the stochastic integral we get

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

where in the last inequality we have used the uniform in θ sub-polynomial growth of f' and b, sub-linear growth of σ and Lemma 1(3). Therefore

$$E\sum_{i=1}^n E\left[\left(e_i^n\right)^2 | \mathcal{F}_{t_{i-1}}\right] \leq C\Delta_n^2 \to 0 \quad \text{when} \quad n\to\infty.$$

We conclude, using Lemma 9 in Genon-Catalot and Jacod [1993], that $\forall\theta\in\Theta$,

$$\frac{1}{\sqrt{n\Delta_n}} a_n(\theta) = \sum_{i=1}^n e_i^n \xrightarrow{P} 0. \tag{102}$$

Using again uniform in θ sub-polynomial growth of b, f', f'', sub-linearity of σ and (3) of the Lemma 1 we easily see that

$$E \sup_{\theta\in\Theta} |b_n(\theta)| \le Cn\Delta_n^2. \tag{103}$$

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

$$\begin{aligned} & E \sup_{\theta\in\Theta} |c_n(\theta)| \\ & \le \sum_{i=1}^n \int_{t_{i-1}}^{t_i} ds \int_{t_{i-1}}^{s} du \int_{\mathbb{R}\setminus\{0\}} E|b(\theta^\star, X_{t_{i-1}})||f(\theta, X_{u-} + \gamma(X_{u-})z) - f(\theta, X_{u-})|\mu(du, dz) \\ & \le \sum_{i=1}^n \int_{t_{i-1}}^{t_i} ds \int_{t_{i-1}}^{s} du \int_{\mathbb{R}\setminus\{0\}} E|b(\theta^\star, X_{t_{i-1}}) f'(\theta, \tilde{x})\gamma(X_{u-})||z|\nu(dz). \end{aligned} \tag{104}$$

(105)

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 $|\tilde{x}| \le |X_{u-}| + |X_u|$. According to the Assumptions 3 (i), (iii) and the assumption b) of the Lemma, the functions γ, $b(\theta^\star, .)$ and $\sup_\theta |f'(\theta, .)|$ are sub-polynomial, and $\nu(|z|) < \infty$. Therefore, using (3) from Lemma 1 we have

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

This last inequality together with (104) gives

$$E \sup_{\theta\in\Theta} |c_n(\theta)| = O(n\Delta_n^2). \tag{106}$$

From (102), (103) and (106) we conclude that under condition $n\Delta_n^{3-\varepsilon} \to 0$,

$$\frac{1}{\sqrt{n\Delta_n}} A_{n,3}(\theta) \xrightarrow{P} 0. \tag{107}$$

Finally, the previous display together with (99) and (101) proves (ii) of the lemma. To prove the claim (i) we will again use the decomposition of the difference given by (95).

Using the same arguments as in (98) and Lemma 15 (1), we get for some $p > 1$, $C > 0$ and $\tilde{x}$ between X_s and $X_{t_{i-1}}$:

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

Hence

$$\frac{1}{n\Delta_n} \sup_{\theta\in\Theta} |A_{n,3}(\theta)| \xrightarrow{L^1} 0. \tag{108}$$

The bound (100) gives

$$\frac{1}{n\Delta_n}\sup_{\theta\in\Theta}|A_{n,2}(\theta)| \xrightarrow{L^1} 0. \tag{109}$$

From (99) 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 θ. Denote $\phi:[0,t_n]\to[0,t_n]$, $\phi(s)=t_{i-1}$ if $t_{i-1}\le s<t_i, i=0,\dots,n-1$, 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 σ and the boundedness of moments of X given by (3) of Lemma 1 we find that for any $p\ge 2$ and some $C>0$,

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

Choosing $p>\frac{d}{\kappa}$ and using the Theorem 20 in the Appendix of I. Ibragimov [2013] we obtain

$$\frac{1}{n\Delta_n}\sup_{\theta\in\Theta}|A_{n,1}(\theta)|\xrightarrow{P} 0$$

and the statement (i) follows.

□

8. Auxiliary results

In this section we gather some auxiliary results that are frequently used in our proofs. Furthermore, we give a proof of the ergodicity results of Lemma 1. We start by some moment inequalities for jump diffusions and their continuous martingale part.

Lemma 15. *Let X satisfy Assumption 1. Then for all $t>s$,*

(1) $\forall p\ge 2$,
$$E[|X_t-X_s|^p]^{1/p}\le C|t-s|^{1/p}.$$
(2) *Let $\mathcal{F}_s=\sigma\{X_u, 0\le u\le s\}$. Then for $p\ge 2$, $p\in\mathbb{N}$,*
$$E[|X_t-X_s|^p|\mathcal{F}_s]\le|t-s|(1+|X_s|^p).$$
(3) $\forall p>1$,
$$E\left[|X_t^c-X_s^c|^p\right]^{1/p}\le C|t-s|^{1/2}.$$

Proof. The first claim follows easily from the two lemmas and Theorem 66 on p. 339 in Protter [2004]. The second claim follows from Proposition 3.1 in Shimizu and Yoshida [2006] and the third from the first two lemmas on p.339 in Protter [2004]. □

Lemma 16. *Under assumptions 1 to 4, we have for some $C>0$,*

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

Proof. We need to introduce some notations. For $z>0$, we define $U_z = \int_{t_{i-1}}^{t_i} \int_{|y|\geq 1/z} \mu(ds,dy)$ the number of jumps of (X_s), $s\in(t_{i-1},t_i]$, with a size greater than $1/z$, and we set $U_0=0$. It is clear that $(U_z)_{z\geq 0}$ is a process whose increments are independent and distributed with Poisson laws. Hence, it is a Poisson process, and by a simple computation we can show that it has a jump intensity equal to $(t_i-t_{i-1})z^{-2}(\nu(z^{-1})+\nu(-z^{-1}))$, where $\nu(z)=\nu(dz)/dz$ exists by Assumption 4 (iii).

We define the filtration generated by the process $(U_z)_{z\geq 0}$, by setting for all $z\geq 0$, $\mathcal{G}_z = \sigma\{U_y;\ y\leq z\}$. We note Z_1^* the first jump time of the process U, which is a stopping time. By construction, we have that $1/Z_1^*$ is the size of the biggest jumps of the Lévy process L on $(t_{i-1},t_i]$, or with the notations of Lemma 9 that, $1/Z_1^* = |\Delta L_{T_i^*}|$, where $|\Delta L_{T_i^*}| = \max\{|\Delta L_s|;\ s\in(t_{i-1};t_i]\}$.

Moreover, we can write

$$\sum_{t_{i-1}<s\leq t_i; s\neq T_i^*} |\Delta L_s| = \int_{t_{i-1}}^{t_i} \int_{|y|<1/Z_1^*} |y| \mu(ds,dy) = \int_{(Z_1^*,\infty)} \frac{1}{z} dU_z,$$

where we have used that $\Delta L_{T_i^*}$ is the only jump with the maximal size $1/Z_1^*$. Hence, we have

$$\begin{aligned}
P((N_n^i)^c \cap (A_n^i)^c) &= P\left(|\Delta L_{T_i^*}| > \frac{3v_n}{\gamma_{min}};\quad \sum_{t_{i-1}<s\leq t_i; s\neq T_i^*} |\Delta L_s| > \frac{v_n}{\gamma_{max}}\right) \\
&= P\left((Z_1^*)^{-1} > \frac{3v_n}{\gamma_{min}};\quad \int_{(Z_1^*,\infty)} z^{-1} dU_z > \frac{v_n}{\gamma_{max}}\right) \\
&= E\left[1_{\{(Z_1^*)^{-1} > \frac{3v_n}{\gamma_{min}}\}} P\left(\int_{(Z_1^*,\infty)} z^{-1} dU_z > \frac{v_n}{\gamma_{max}} \mid \mathcal{G}_{Z_1^*}\right)\right] \\
&\leq \frac{\gamma_{max}}{v_n} E\left[1_{\{(Z_1^*)^{-1} > \frac{3v_n}{\gamma_{min}}\}} E\left(\int_{(Z_1^*,\infty)} z^{-1} dU_z \mid \mathcal{G}_{Z_1^*}\right)\right],
\end{aligned}$$

where we have used the Markov inequality in the last line. Using now that $(U_z)_{z\geq 0}$ is a Poisson process with an explicit jump intensity $\overline{U}(z) := (t_i-t_{i-1})z^{-2}(\nu(z^{-1})+\nu(-z^{-1}))$, we deduce,

$$P((N_n^i)^c \cap (A_n^i)^c) \leq \frac{\gamma_{max}}{v_n} E\left[1_{\{(Z_1^*)^{-1} > \frac{3v_n}{\gamma_{min}}\}} E\left(\int_{(Z_1^*,\infty)} z^{-1} \overline{U}(z) dz \mid \mathcal{G}_{Z_1^*}\right)\right].$$

But, by a simple change of variable, $\int_{(Z_1^*,\infty)} z^{-1}\overline{U}(z)dz = (t_i - t_{i-1}) \int_{|y|<1/Z_1^*} |y|\nu(y)dy \le \Delta_n \int_{\mathbb{R}} |y|\nu(y)\mathrm{d}y$. We conclude

$$\begin{aligned}
P((N_n^i)^c \cap (A_n^i)^c) &\le \frac{\gamma_{max}}{v_n}\Delta_n \left(\int_{\mathbb{R}} |y|\nu(y)\mathrm{d}y\right) P\left[(Z_1^*)^{-1} > \frac{3v_n}{\gamma_{min}}\right] \\
&\le C\frac{\Delta_n}{v_n} P\left(\mu((t_{i-1}, t_i] \times [(-\infty, -\frac{3v_n}{\gamma_{min}}) \cup (\frac{3v_n}{\gamma_{min}}, +\infty)]) \ge 1\right) \\
&\le C\frac{\Delta_n^2}{v_n} \int_{|z|>\frac{3v_n}{\gamma_{min}}} \nu(dz),
\end{aligned}$$

where $C > 0$. The lemma is proved. □

Proposition 17. *Under Assumptions 1 to 4, the Assumption 6 is equivalent to the condition*

$$\forall (\theta, \theta') \in \Theta^2, \quad \textit{such that} \quad \theta \neq \theta', \quad b(\theta, .) \neq b(\theta', .).$$

Proof. It is sufficient to show that if $\mathcal{O}$ is some non empty, open set, then $\pi^\theta(\mathcal{O}) > 0$. It is proved in Masuda [2007] (see equation (13) p.43) that for all $\Delta > 0$, $x \in \mathbb{R}$, and $\mathcal{O}$ non empty, open set, $P(X_\Delta^\theta \in \mathcal{O} \mid X_0^\theta = x) > 0$. From this, we deduce that

$$\pi^\theta(\mathcal{O}) = \int_{\mathbb{R}} P(X_\Delta^\theta \in \mathcal{O} \mid X_0^\theta = x) d\pi^\theta(x) > 0.$$

□

We conclude this section with a proof of the ergodicity results and moment bounds of Lemma 1. The proof is based on Masuda [2007].

Proof of Lemma 1. Let $q > 2$, q even and $f^\star(x) = |x|^q$. We show that $f^\star$ satisfies the drift condition

$$\mathcal{A}f^\star \le -c_1 f^\star + c_2,$$

where $c_1 > 0, c_2 > 0$. Denote

$$\mathcal{G}f(x) = \frac{1}{2}\sigma^2(x) f''(x) + b(\theta, x) f'(x),$$

$$\mathcal{J}f(x) = \int_{\mathbb{R}} (f(x + z\gamma(x)) - f(x))\nu(dz).$$

for any f such that the two previous expressions are defined and decompose

$$\mathcal{A} = \mathcal{G} + \mathcal{J}.$$

Using Taylor's formula together with Assumptions 3 (iii) and 4 (ii) we can write

$$|\mathcal{J}f^\star(x)| \le \int_{\mathbb{R}} |z\gamma(x)| \sup_{u\in[x, x+z\gamma(x)]} |f^{\star'}(u)|\nu(dz) \le C\gamma(x)|x|^{q-1} \int_{\mathbb{R}} |z|(1+|z|)^{q-1}\nu(dz) = o(|x|^q)$$

as $x \to \infty$. Using Assumption 3 (ii) and (iv) we get

$$\mathcal{G}f^\star(x) = \frac{1}{2}\sigma^2(x) q(q-1)x^{q-2} + b(\theta, x) x q x^{q-2} \le -C|x|^2 q x^{q-2} + o(|x|^q) \le -Cqf^\star(x) + o(|x|^q),$$

for some $C > 0$. As $\mathcal{A}f^\star(x)$ is locally bounded, using two previous displays we can choose $c_2 > 0$ and $c_1 > 0$ such that for all $x \in \mathbb{R}$,

$$\mathcal{A}f^\star(x) \le -c_1 f^\star(x) + c_2.$$

Hence, Assumption $3^\star$ from Masuda [2007] holds and using Theorem 2.2 from Masuda [2007] we get then

$$\sup_{s\geq 0} \mathbf{E}[|X_s^\theta|^q] < \infty \tag{110}$$

and using Fatou's lemma results in

$$\sup_{s\geq 0} \mathbf{E}[|X_{s_-}^\theta|^q] < \infty.$$

Hence we proved the assertion (3). Using Assumption 2 and the Theorem 2.1 from Masuda [2007] we get for all $\theta \in \Theta$ that X^θ admits the unique invariant distribution π^θ, $f^\star \in \mathcal{L}^1(\pi^\theta)$ and the ergodic theorem holds. We proved (1) and (2). We continue with the proof of (4). Using ergodic theorem, for all $q > 0$,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^\infty |X_s^\theta|^q ds = \pi^\theta(|x|^q), \quad P - a.s.$$

Moreover, using Jensen's inequality and the bound (110) we get the uniform integrability of the family $\{\frac{1}{t}\int_0^t |X_s^\theta|^q ds,\ t > 0\}$:

$$E\left(\frac{1}{t}\int_0^t |X_s^\theta|^q ds\right)^{1+\varepsilon} \leq \frac{1}{t}\int_0^t [E|X_s^\theta|^{q(1+\varepsilon)}]ds \leq C,$$

where $C > 0$, and hence

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

□

Proof of Lemma 11. Let us first prove (i). Using Lemma 15 (1), with some $\tilde{x}$ between $X_{t_{i-1}}$ and X_s in the third line below we obtain:

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

We now prove (ii). We find that

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

and it is then apparent that this term can be treated exactly as the term $A_{n,3}(\theta)$ given by the equation (97). Hence, from (107) (which requires the condition $n\Delta_n 3 - \varepsilon \to 0$, we have the result.

□

References

David Applebaum. *Lévy processes and stochastic calculus. 2nd ed.* Cambridge Studies in Advanced Mathematics 116. Cambridge University Press., 2009. 2.1

Ole E. Barndorff-Nielsen and Neil Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc., Ser. B, Stat. Methodol.*, 63(2):167–241, 2001. 1, 5.1

Markus Bibinger and Lars Winkelmann. Econometrics of co-jumps in high-frequency data with noise. *Journal of Econometrics*, 184(2):361 – 378, 2015. ISSN 0304-4076. doi: http://dx.doi.org/10.1016/j.jeconom.2014.10.004. 1

Rama Cont and Peter Tankov. *Financial modelling with jump processes.* Chapman & Hall, 2004. 5

John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, (53):363–384, 1985. 5.1

Susanne Ditlevsen and Priscilla Greenwood. The morris–lecar neuron model embeds a leaky integrate-and-fire model. *Journal of Mathematical Biology*, 67(2):239–259, 2013. ISSN 0303-6812. doi: 10.1007/s00285-012-0552-7. URL `http://dx.doi.org/10.1007/s00285-012-0552-7`. 1

D. Florens-Zimrou. Approximate discrete-time schemes for statistics of diffusion processes,. *Statistics*, 20:547–557, 1989. 1

Valentine Genon-Catalot and Jean Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 29(1):119–151, 1993. 7, 7, 7

James E. Hutton and Paul . Nelson. Interchanging the order of differentiation and stochastic integration. *Stochastic Processes and their Applications*, 18(2):371–377, 1984. 6.1

R. Has'minskii I. Ibragimov. *Statistical estimation:asymptotic theory.* 2013. 7

Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes. 2nd ed.* Grundlehren der Mathematischen Wissenschaften. 288. Berlin: Springer, 2003. 3.1

Mathieu Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24(2):211–229, 1997. ISSN 1467-9469. doi: 10.1111/1467-9469.00059. URL `http://dx.doi.org/10.1111/1467-9469.00059`. 1

S. G. Kou. A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101, 2002. doi: 10.1287/mnsc.48.8.1086.166. URL `http://dx.doi.org/10.1287/mnsc.48.8.1086.166`. 1

Uwe Küchler and Michael Sørensen. A note on limit theorems for multivariate martingales. *Bernoulli*, 5(3):483–493, 1999. URL `http://dx.doi.org/10.2307/3318713`. 6.1, 6.2

D. Loukianova and O. Loukianov. Uniform law of large numbers and consistency of estimators for Harris diffusions. *Statist. Probab. Lett.*, 74(4):347–355, 2005. ISSN 0167-7152. doi: 10.1016/j.spl.2005.04.056. URL `http://dx.doi.org/10.1016/j.spl.2005.04.056`. 6.1, 6.1, 6.1

Hilmar Mai. Efficient maximum likelihood estimation for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 20(2):919–957, 05 2014. doi: 10.3150/13-BEJ510. URL `http://dx.doi.org/10.3150/13-BEJ510`. 1

Cecilia Mancini. The speed of convergence of the threshold estimator of integrated variance. *Stochastic Process. Appl.*, 121(4):845–855, 2011. 1

Hiroki Masuda. Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.*, 117(1):35–56, 2007. 2.1, 2.2, 8, 8

Hiroki Masuda. Erratum to :"ergodicity and exponential β-mixing bound for multidimensional diffusions with jumps". *Stochastic Processes and their Applications*, 119:676–678, 2009. 2.1

Hiroki Masuda. Convergence of gaussian quasi-likelihood random fields for ergodic lévy driven sde observed at high frequency. *Ann. Statist.*, 41(3):1593–1641, 06 2013. doi: 10.1214/13-AOS1121.

URL http://dx.doi.org/10.1214/13-AOS1121. 1

Robert Merton. Option pricing when underlying stock returns are discontinu- ous. *J. Financial Economics*, 3:125–144, 1976. 1

Philip E. Protter. *Stochastic integration and differential equations. 2nd ed.* Applications of Mathematics 21. Berlin: Springer, 2004. 15

Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1994. ISBN 3-540-57622-3. 6.3

Yasutaka Shimizu. M-estimation for doscretely observed ergodic diffusion processes with infinetely many jumps. *Stat. Inference Stoch. Process.*, 9:179–225, 2006. 1

Yasutaka Shimizu and Nakahiro Yoshida. Estimation of parameters for diffusion processes with jumps from discrete observations. *Statistical Inference for Stochastic Processes*, 9(3):227–277, 2006. ISSN 1387-0874. doi: 10.1007/s11203-005-8114-x. URL http://dx.doi.org/10.1007/s11203-005-8114-x. 1, 15

Ngoc Khue Tran. *LAN property for jump diffusion processes with discrete observations via Malliavin calculus.* PhD thesis, University Paris 13, 2014. 1

Aad W. Van der Vaart. *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press., 1998. 6.1, 6.3

Nakahiro Yoshida. Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41(2):220 – 242, 1992. 1

Université d'Evry Val d'Essonne, 91037 Évry Cedex, France
E-mail address: arnaud.gloter@univ-evry.fr

Université d'Evry Val d'Essonne, 91037 Évry Cedex, France
E-mail address: dasha.loukianova@maths.univ-evry.fr

Centre de Recherche en Economie et Statistique, ENSAE-ParisTech, 92245 Malakoff, France
E-mail address: hilmar.mai@ensae.fr