

Joint estimation for volatility and drift parameters of ergodic jump diffusion processes via contrast function

Chiara Amorino, Arnaud Gloter

▶ **To cite this version:**

Chiara Amorino, Arnaud Gloter. Joint estimation for volatility and drift parameters of ergodic jump diffusion processes via contrast function. 2020. hal-02331156v2

HAL Id: hal-02331156
https://hal.science/hal-02331156v2

Preprint submitted on 24 Nov 2020

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

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

Joint estimation for volatility and drift parameters of ergodic jump diffusion processes via contrast function.

Chiara Amorino*, Arnaud Gloter*

November 24, 2020

Abstract

In this paper we consider an ergodic diffusion process with jumps whose drift coefficient depends on μ and volatility coefficient depends on σ, two unknown parameters. We suppose that the process is discretely observed at the instants $(t_i^n)_{i=0,\dots,n}$ with $\Delta_n = \sup_{i=0,\dots,n-1}(t_{i+1}^n - t_i^n) \to 0$. We introduce an estimator of $\theta := (\mu, \sigma)$, based on a contrast function, which is asymptotically gaussian without requiring any conditions on the rate at which $\Delta_n \to 0$, assuming a finite jump activity. This extends earlier results where a condition on the step discretization was needed (see [15],[36]) or where only the estimation of the drift parameter was considered (see [2]). In general situations, our contrast function is not explicit and in practise one has to resort to some approximation. We propose explicit approximations of the contrast function, such that the estimation of θ is feasible under the condition that $n\Delta_n^k \to 0$ where $k > 0$ can be arbitrarily large. This extends the results obtained by Kessler [24] in the case of continuous processes.

Drift estimation, volatility estimation, ergodic properties, high frequency data, Lévy-driven SDE, thresholding methods.

1 Introduction

Recently, diffusion processes with jumps are becoming powerful tools to model various stochastic phenomena in many areas, for example, physics, biology, medical sciences, social sciences, economics, and so on. In finance, jump-processes were introduced to model the dynamic of exchange rates ([6]), asset prices ([29],[25]), or volatility processes ([5],[11]). Utilization of jump-processes in neuroscience, instead, can be found for instance in [9]. Therefore, inference problems for such models from various types of data should be studied, in particular, inference from discrete observation should be desired since the actual data may be obtained discretely.
In this work, our aim is to estimate jointly the drift and the volatility parameter $(\mu, \sigma) =: \theta$ from a discrete sampling of the process X^θ solution to

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

where W is a one dimensional Brownian motion and $\tilde{\mu}$ a compensated Poisson random measure, with a finite jump activity. We assume that the process is sampled at the times $(t_i^n)_{i=0,\dots,n}$ where the sampling step $\Delta_n := \sup_{i=0,\dots,n-1} t_{i+1}^n - t_i^n$ goes to zero. A crucial point for applications in the high frequency setting is to impose minimal conditions on the sampling step size. This will be one of our main objectives in this paper, for the joint estimation of μ and σ.
It is known that, as a consequence of the presence of a Gaussian component, it is impossible to estimate the drift parameter on a finite horizon time; we therefore assume that $t_n^n \to \infty$ and we suppose to have an ergodic process X^θ.

The topic of high frequency estimation for discretely observed diffusions in the case without jumps is well developed, by now. Florens-Zmirou has introduced, in [13], an estimator for both the drift and the diffusion parameters under the fast sampling assumption $n\Delta_n^2 \to 0$. Yoshida [37] has then suggested a correction of the contrast function of [13] that releases the condition on the step discretization to $n\Delta_n^3 \to 0$. In Kessler [24], the author proposes an explicit modification of the Euler scheme contrast that allows him to build an estimator which is asymptotically normal under the condition $n\Delta_n^k \to 0$ where $k \geq 2$ is arbitrarily large. The result found by Kessler, therefore, holds for any arbitrarily slow polynomial decay to zero of the sampling step.

*Laboratoire de Mathématiques et Modélisation d'Evry, CNRS, Univ Evry, Université Paris-Saclay, 91037, Evry, France.

When a jump component is added, less results are known. Shimizu [34] proposes parametric estimation of drift, diffusion and jump coefficients showing the asymptotic normality of the estimators under some explicit conditions relating the sampling step and the jump intensity of the process; such conditions on Δ_n are more restrictive as the intensity of jumps near zero is high. In the situation where the jump intensity is finite, the conditions of [34] reduces to $n\Delta_n^2 \to 0$. In [15], the condition on the sampling step is relaxed to $n\Delta_n^3 \to 0$, when one estimates the drift parameter only. In [30] a jump-filtering technique similar to one used in [15] is employed in order to derive a nonparametric estimator for the drift which is robust to symmetric jumps of infinite variance and infinite variation, and which attains the same asymptotic variance as for a continuous diffusion process. Also in [2] only the estimation of the drift parameter is studied. In such a case, the sampling step $(t_i^n)_{i=0,\dots,n}$ can be irregular, no condition on the rate at which $\Delta_n \to 0$ is needed and the assumption that the jumps of the process are summable, present in [15], is suppressed.

In this paper, we consider the joint estimation of the drift and the diffusion parameters with a jump intensity which is finite. Since for the applications it is important that assumptions on the rate at which Δ_n should tend to zero are less stringent as possible, our aim is to weaken the conditions on the decay of the sampling step in a way comparable to Kessler's work [24], but in the framework of jump-diffusion processes. We therefore want to extend [2] looking for the same results, but for the joint estimation of the drift and the diffusion parameters instead of focusing on the drift parameter only.
The joint estimation of the two parameters introduces some significant difficulties: since the drift and the volatility parameters are not estimated at the same rate, we have to deal with asymptotic properties in two different regimes.
Compared to previous results in which the parameters are estimated jointly (see [34]), we show that it is possible to remove any condition on the rate at which Δ_n has to go to zero.
Moreover, we consider a discretization step which is not uniform. This case, to our knowledge, has never been studied before for the joint estimation of the drift and the volatility of a diffusion with jumps.

A natural approach to estimate the unknown parameters would be to use a maximum likelihood estimation, but the likelihood function based on the discrete sample is not tractable in this setting, since it depends on the transition densities of X which are not explicitly known.
To overcome this difficulty several methods have been developed. For instance, in [1] and [26] closed form expansions of the transition density of jump-diffusions are studied while in [22] the asymptotic behaviour of estimating functions is considered in the high frequency observation framework. They give condition to ensure the rate optimality and the efficiency.

Considering again the case of high frequency observation, a widely-used method is to consider pseudo-likelihood function, for instance based on the high frequency approximation of the dynamic of the process by the dynamic of the Euler scheme. This leads to explicit contrast functions with Gaussian structures (see e.g. [36],[34],[28]).

In Kessler's paper the idea is to replace, in the Euler scheme contrast function, the contribution of the drift and the diffusion by two quantities m and m_2 (or their explicit approximations with arbitrarily high order when $\Delta_n \to 0$); with

$$m(\mu,\sigma,x) := \mathbb{E}[X^\theta_{t_{i+1}} | X^\theta_{t_i} = x] \quad \text{and} \tag{1}$$

$$m_2(\mu,\sigma,x) := \mathbb{E}[(X^\theta_{t_{i+1}} - m(\mu,\sigma,X^\theta_{t_i}))^2 | X^\theta_{t_i} = x].$$

In presence of jumps, the contrasts functions in [36] (see also [34], [15]) resort to a filtering procedure in order to suppress the contribution of jumps and recover the continuous part of the process.

The contrast function we introduce is based on both the ideas described here above. Indeed, we define it as

$$U_n(\mu,\sigma) := \sum_{i=0}^{n-1} [\frac{(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2}{m_2(\mu,\sigma,X_{t_i})} + \log(\frac{m_2(\mu,\sigma,X_{t_i})}{\Delta_{n,i}})]\varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}, \tag{2}$$

where the function φ is a smooth version of the indicator function that vanishes when the increments of the data are too large compared to the typical increments of a continuous diffusion process, and thus can be used to filter the contribution of the jumps. The idea is to use the size of $X_{t_{i+1}} - X_{t_i}$ in order to judge if a jump occurred or not in the interval $[t_i, t_{i+1})$. The increment of X with continuous transition could hardly exceed the threshold $\Delta^\beta_{n,i}$, therefore we can judge the existence of a jump in the interval if $|X_{t_{i+1}} - X_{t_i}| > \Delta^\beta_{n,i}$, for $\beta \in (\frac{1}{4}, \frac{1}{2})$.
The last indicator in (2) avoids the possibility that $|X_{t_i}|$ is too big, the constant k is positive and will be chosen later, related to the developments of m and m_2 that are the natural extension to the case with jumps of the quantities proposed in [24]. Indeed, we have defined them as

$$m(\mu,\sigma,x) := \frac{\mathbb{E}[X^\theta_{t_{i+1}} \varphi_{\Delta^\beta_{n,i}}(X^\theta_{t_{i+1}} - X^\theta_{t_i}) | X^\theta_{t_i} = x]}{\mathbb{E}[\varphi_{\Delta^\beta_{n,i}}(X^\theta_{t_{i+1}} - X^\theta_{t_i}) | X^\theta_{t_i} = x]} \quad \text{and}$$

$$m_2(\mu,\sigma,x) := \frac{\mathbb{E}[(X^\theta_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2 \varphi_{\Delta^\beta_{n,i}}(X^\theta_{t_{i+1}} - X^\theta_{t_i})|X^\theta_{t_i} = x]}{\mathbb{E}[\varphi_{\Delta^\beta_{n,i}}(X^\theta_{t_{i+1}} - X^\theta_{t_i})|X^\theta_{t_i} = x]}.$$

The rates for the estimation of the two parameters is not the same, which implies we have to deal with two different scaling of the contrast function, which lead us to the study of the asymptotic properties of the contrast function in two different asymptotic schemes.

The main result of our paper is that the estimator $\hat{\theta}_n := (\hat{\mu}_n, \hat{\sigma}_n)$ associated to the proposed contrast function converges with some explicit asymptotic variances. Comparing to earlier results ([36], [34], [15], [2]), the sampling step $(t^n_i)_{i=0,\dots,n}$ can be irregular, no condition is needed on the rate at which $\Delta_n \to 0$ and the parameters of drift and diffusion are jointly estimated.

Moreover, we provide explicit approximations of m_2 that allows us to circumvent the fact that the contrast function is non explicit (explicit approximations of m are given in [2]). We give an expansion of m_2 exact up to order Δ_n^2, which involves the jump intensity near zero, and is valid for any smooth truncation function φ. With the specific choices of φ being oscillating functions, in particular, we remove the contribution of the jumps and we are able to prove explicit developments of the function m_2 valid up to any order. Together with the approximation of the function m showed in Proposition 2 of [2], this allows us to approximate our contrast function, at arbitrary high order, by a completely explicit one, as it was in the paper by Kessler [24] in the continuous case.
This yields to a consistent and asymptotic normal estimator under the condition $n\Delta_n^k \to 0$, where k is related to the oscillating properties of the function φ. As k can be chosen arbitrarily high, up to a proper choice of φ, our method allows to estimate the drift and the diffusion parameters, under the assumption that the sampling step tends to zero at some polynomial rate.

Furthermore, we implement numerically our main results building two approximations of m and m_2 from which we deduce two approximations of the contrast that we minimize in order to get the joint estimator of the parameters. We compare the estimators we find with the estimator that would result from the choice of an Euler scheme approximation for m and m_2. From our simulations it appears that our joint estimator performs better than the Euler one, especially for the estimation of the parameter σ.

The outline of the paper is the following. In Section 2 we introduce the model and we state the assumptions we need. The Section 3 contains the construction of the estimator and our main results while in Section 4 we explain how to use in practical the contrast function for the joint estimation of the drift and the diffusion parameters, dealing with its approximation. We provide numerical results in Section 5. In Section 6 we discuss about possible forward generalization of the obtained results, while in Section 7 we state useful propositions that we will use repeatedly in the following sections. Section 8 is devoted to the proof of our main results while in Section A.1 of the Appendix we prove the propositions stated in the sixth section. We conclude giving the proofs of some technical results in the Sections A.2–A.3 of the Appendix.

2 Model, assumptions

We want to estimate the unknown parameter $\theta = (\mu, \sigma)$ in the stochastic differential equation with jumps

$$X^\theta_t = X^\theta_0 + \int_0^t b(\mu, X^\theta_s)ds + \int_0^t a(\sigma, X^\theta_s)dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X^\theta_{s^-}) z\tilde{\mu}(ds,dz), \quad t \in \mathbb{R}_+, \tag{3}$$

where θ belongs to $\Theta := \Pi \times \Sigma$, a compact set of $\mathbb{R}^2$; $W = (W_t)_{t\geq 0}$ is a one dimensional Brownian motion and μ is a Poisson random measure associated to the Lévy process $L = (L_t)_{t\geq 0}$ such that $L_t := \int_0^t \int_{\mathbb{R}} z\tilde{\mu}(ds,dz)$. The compensated measure $\tilde{\mu} = \mu - \bar{\mu}$ is defined on $[0,\infty) \times \mathbb{R}$, the compensator is $\bar{\mu}(dt,dz) := F(dz)dt$, where conditions on the Levy measure F will be given later.
We denote by $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space on which W and μ are defined and we assume that the initial condition X^θ_0, W and L are independent.

2.1 Assumptions

We suppose that the functions $b : \Pi \times \mathbb{R} \to \mathbb{R}$, $a : \Sigma \times \mathbb{R} \to \mathbb{R}$ and $\gamma : \mathbb{R} \to \mathbb{R}$ satisfy the following assumptions:

A1: *The functions $\gamma(x)$, $b(\mu,x)$ for all $\mu \in \Pi$ and $a(\sigma,x)$ for all $\sigma \in \Sigma$ are globally Lipschitz. Moreover, the Lipschitz constants of b and a are uniformly bounded on Π and Σ, respectively.*

Under Assumption 1 the equation (3) admits a unique non-explosive càdlàg adapted solution possessing the strong Markov property, cf [4] (Theorems 6.2.9. and 6.4.6.).
The next assumption was used in [27] to prove the irreducibility of the process X^θ.

A2: *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 $z = z(x,U) \in supp(F)$ such that $\gamma(x)z \in U$.*

Assumption 2 ensures, together with the Assumption 3 below, the existence of unique invariant distribution π^θ, as well as the ergodicity of the process X^θ.

A3 (Ergodicity): *(i) For all $q > 0$, $\int_{|z|>1} |z|^q F(z)dz < \infty$.*
(ii) For all $\mu \in \Pi$ there exists $C > 0$ such that $xb(\mu, x) \leq -C|x|^2$, if $|x| \to \infty$.
(iii) $|\gamma(x)|/|x| \to 0$ as $|x| \to \infty$.
(iv) For all $\sigma \in \Sigma$ we have $|a(\sigma, x)|/|x| \to 0$ as $|x| \to \infty$.
(v) $\forall \theta \in \Theta$, $\forall q > 0$ we have $\mathbb{E}|X_0^\theta|^q < \infty$.

A4 (Jumps): *1. The jump coefficient γ is bounded from below, that is $\inf_{x\in\mathbb{R}} |\gamma(x)| := \gamma_{min} > 0$.*
2. The Lévy measure F is absolutely continuous with respect to the Lebesgue measure and we denote $F(z) = \frac{F(dz)}{dz}$.
3. F is such that $F(z) = \lambda F_0(z)$ and $\int_{\mathbb{R}} F_0(z)dz = 1$.

Assumption 4.1 is useful to compare size of jumps of X and L. The Assumption 5 ensures the existence of the contrast function we will define in next section.

A5 (Non-degeneracy): *There exists some $c > 0$, such that $\inf_{x,\sigma} a^2(\sigma, x) \geq c > 0$.*

From now on we denote the true parameter value by θ_0, an interior point of the parameter space Θ that we want to estimate. We shorten X for X^{θ_0}.

We will use some moment inequalities for jump diffusions, gathered in the following lemma that follows from Theorem 66 of [32] and Proposition 3.1 in [36].

Lemma 1. *Let X satisfies Assumptions 1-4. Let $L_t := \int_0^t \int_{\mathbb{R}} z\tilde{\mu}(ds, dz)$ and let $\mathcal{F}_s := \sigma\{(W_u)_{0<u\leq s}, (L_u)_{0<u\leq s}, X_0\}$. Then, for all $t > s > 0$,*
1) for all $p \geq 2$, $\mathbb{E}[|X_t - X_s|^p]^{\frac{1}{p}} \leq c|t-s|^{\frac{1}{p}}$,
2) for all $p \geq 2$, $p \in \mathbb{N}$, $\mathbb{E}[|X_t - X_s|^p|\mathcal{F}_s] \leq c|t-s|(1 + |X_s|^p)$,
3) for all $p \geq 2$, $p \in \mathbb{N}$, $\sup_{h\in[0,1]} \mathbb{E}[|X_{s+h}|^p|\mathcal{F}_s] \leq c(1 + |X_s|^p)$.

An important role is played by ergodic properties of solution of equation (3)
The following Lemma states that Assumptions $1-4$ are sufficient for the existence of an invariant measure π^θ such that an ergodic theorem holds and moments of all order exist.

Lemma 2. *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 $g : \mathbb{R} \to \mathbb{R}$ s. t. $\pi^\theta(g) < \infty$, 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 $q > 0$, $\sup_{t\geq 0} \mathbb{E}[|X_t^\theta|^q] < \infty$.

A proof is in [15] (Section 8 of Supplement), it relies mainly on results of [27].

A6 (Identifiability): *For all μ_1, μ_2 in Π, $\mu_1 = \mu_2$ if and only if $b(\mu_1, x) = b(\mu_2, x)$ for almost all x. Moreover, $\forall \sigma_1, \sigma_2$ in Σ, $\sigma_1 = \sigma_2$ if and only if $a(\sigma_1, x) = a(\sigma_2, x)$ for almost all x.*

A7: *1. The derivatives $\frac{\partial^{k_1+k_2} b}{\partial x^{k_1} \partial \theta^{k_2}}$, with $k_1 + k_2 \leq 4$ and $k_2 \leq 3$, exist and they are bounded if $k_1 \geq 1$. If $k_1 = 0$, for each $k_2 \leq 3$ they have polynomial growth.*
2. The derivatives $\frac{\partial^{k_1+k_2} a}{\partial x^{k_1} \partial \theta^{k_2}}$, with $k_1 + k_2 \leq 4$ and $k_2 \leq 3$, exist and they are bounded if $k_1 \geq 1$. If $k_1 = 0$, for each $k_2 \leq 3$ they have polynomial growth.
3. The derivatives $\gamma^{(k)}(x)$ exist and they are bounded for each $1 \leq k \leq 4$.

A8: *Let B be* $\begin{pmatrix} -2\int_{\mathbb{R}} (\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2 \pi(dx) & 0 \\ 0 & 4\int_{\mathbb{R}} (\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2 \pi(dx) \end{pmatrix}$, *then $det(B) \neq 0$.*

3 Construction of the estimator and main results

Now we present a contrast function for estimating parameters.

3.1 Construction of contrast function.

Suppose that we observe a finite sample $X_{t_0}, ..., X_{t_n}$ with $0 = t_0 \leq t_1 \leq ... \leq t_n =: T_n$, where X is the solution to (3) with $\theta = \theta_0$. Every observation time point depends also on n, but to simplify the notation we suppress this index. We will be working in a high-frequency setting, i.e. $\Delta_n := \sup_{i=0,...,n-1} \Delta_{n,i} \longrightarrow 0$ for $n \to \infty$, with $\Delta_{n,i} := (t_{i+1} - t_i)$. We assume that $\lim_{n\to\infty} T_n = \infty$.
In the sequel we will always suppose that the following assumption on the step discretization holds true.
$A_{\textbf{Step}}$**:** there exist two constants c_1, c_2 such that $c_2 < \frac{\Delta_n}{\Delta_{min}} < c_1$, where we have denoted Δ_{min} as $\min_{i=0,...,n-1} \Delta_{n,i}$.
We introduce a version of the gaussian quasi -likelihood in which we have filtered the contribution of the jumps. As in our framework the data is observed discretely, to formulate a criterion to decide if a jump occurred in a particular interval or not we can only use the increment of our process. Such a criterion should depends on n and should be more and more accurate as n tends to infinity. This leads to the following contrast function:

Definition 1. *For $\beta \in (\frac{1}{4}, \frac{1}{2})$ we define the contrast function $U_n(\mu, \sigma)$ as*

$$U_n(\mu,\sigma) := \sum_{i=0}^{n-1} [\frac{(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2}{m_2(\mu,\sigma,X_{t_i})} + \log(\frac{m_2(\mu,\sigma,X_{t_i})}{\Delta_{n,i}})]\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}}, \tag{4}$$

with

$$m(\mu,\sigma,x) := \frac{\mathbb{E}[X^\theta_{t_{i+1}} \varphi_{\Delta_{n,i}^\beta}(X^\theta_{t_{i+1}} - X^\theta_{t_i}) | X^\theta_{t_i} = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X^\theta_{t_{i+1}} - X^\theta_{t_i}) | X^\theta_{t_i} = x]}; \tag{5}$$

$$m_2(\mu,\sigma,x) := \frac{\mathbb{E}[(X^\theta_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2 \varphi_{\Delta_{n,i}^\beta}(X^\theta_{t_{i+1}} - X^\theta_{t_i}) | X^\theta_{t_i} = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X^\theta_{t_{i+1}} - X^\theta_{t_i}) | X^\theta_{t_i} = x]} \tag{6}$$

and

$$\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) = \varphi(\frac{X_{t_{i+1}} - X_{t_i}}{\Delta_{n,i}^\beta}).$$

The function φ is a smooth version of the indicator function, such that $\varphi(\zeta) = 0$ for each ζ, with $|\zeta| \geq 2$ and $\varphi(\zeta) = 1$ for each ζ, with $|\zeta| \leq 1$.
The last indicator aims to avoid the possibility that $|X_{t_i}|$ is big. The constant k is positive and it will be chosen later, related to the development of both m and m_2.
Moreover we remark that m and m_2 depend also on t_i and t_{i+1}. By the homogeneity of the equation they actually depend on the difference $t_{i+1} - t_i$ but we omit such a dependence in the notation of the two functions here above to make the reading easier.
We define an estimator $\hat{\theta}_n$ of θ_0 as

$$\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n) \in \arg\min_{(\mu,\sigma)\in\Theta} U_n(\mu,\sigma). \tag{7}$$

The idea is to use the size of the increment of the process $X_{t_{i+1}} - X_{t_i}$ in order to judge if a jump occurred or not in the interval $[t_i, t_{i+1})$. As it is hard for the increment of X with continuous transition to overcome the threshold $\Delta_{n,i}^\beta$ for $\beta \leq \frac{1}{2}$, we can assert the presence of a jump in $[t_i, t_{i+1})$ if $|X_{t_{i+1}} - X_{t_i}| > \Delta_{n,i}^\beta$.
It is worth noting that if β is too large, and therefore $\Delta_{n,i}^\beta$ is too small, we can't ignore the probability to have continuous diffusion reaching the threshold $\Delta_{n,i}^\beta$. On the other side, if β is too small and therefore $\Delta_{n,i}^\beta$ is too big, it is possible to have an increment smaller than $\Delta_{n,i}^\beta$ even if a jump occurs in the considered interval. That's the reason why the threshold β has to be chosen with great care.
In the definition of the contrast function we have taken $\beta > \frac{1}{4}$ because, in Lemma 3 and Proposition 6 below (and so, as a consequence, in the majority of the theorems of this work), such a technical condition on β is required.
We observe that, in general, there is no closed expression for m and m_2, hence the contrast is not explicit. However, it is proved in [2] an explicit development of m in the case where the intensity is finite and in this work we provide as well an explicit development of m_2 that lead us to an explicit version of our contrast function.

3.2 Main results

Before stating our main results, we need some further notations and assumptions. We introduce the function R defined as follows: for $\delta \geq 0$, we will denote $R(\theta, \Delta_{n,i}^{\delta}, x)$ for any function $R(\theta, \Delta_{n,i}^{\delta}, x) = R_{i,n}(\theta, x)$, where $R_{i,n} : \Theta \times \mathbb{R} \longrightarrow \mathbb{R}$, $(\theta, x) \mapsto R_{i,n}(\theta, x)$ is such that

$$\exists c > 0 \qquad |R_{i,n}(\theta, x)| \leq c(1 + |x|^c)\Delta_{n,i}^{\delta} \tag{8}$$

uniformly in θ and with c independent of i, n.
The functions R represent the term of rest and have the following useful property, consequence of the just given definition:

$$R(\theta, \Delta_{n,i}^{\delta}, x) = \Delta_{n,i}^{\delta} R(\theta, \Delta_{n,i}^{0}, x). \tag{9}$$

We point out that it does not involve the linearity of R, since the functions R on the left and on the right side are not necessarily the same but only two functions on which the control (8) holds with $\Delta_{n,i}^{\delta}$ and $\Delta_{n,i}^{0}$, respectively.

In the sequel, we will need a development for the function m_2. We will assume that such a development exists, as stated in the next assumption:

Ad: There exist three functions $r(\mu, \sigma, x)$, $r(x)$, $R(\theta, 1, x)$ and $\delta_1, \delta_2 > 0$ and $k_0 > 0$ such that, for $|x| \leq \Delta_{n,i}^{-k_0}$,

$$m_2(\mu, \sigma, x) = \Delta_{n,i} a^2(x, \sigma)(1 + \Delta_{n,i} r(\mu, \sigma, x)) + \Delta_{n,i}^{1+\delta_1} r(x) + \Delta_{n,i}^{2+\delta_2} R(\theta, 1, x), \tag{10}$$

where $r(\mu, \sigma, x)$ and $r(x)$ are particular functions $R(\theta, 1, x)$, that turns out from the development of m_2, and the function $r(x)$ does not depend on θ. Moreover, the order of such functions does not change by deriving them with respect to both the parameters, that is for $\vartheta = \mu$ and $\vartheta = \sigma$, $|\frac{\partial}{\partial \vartheta} r(\mu, \sigma, x)| \leq c(1+|x|^c)$ and $|\frac{\partial}{\partial \vartheta} R(\theta, 1, x)| \leq c(1 + |x|^c)$.

Assumption Ad is not restrictive. Examples of frameworks in which Ad holds are introduced in Propositions 2 and 4, that will be stated in the next section and proven in the appendix. Let us stress that it is crucial for the proof of the consistency of the estimator that the second main term of the expansion (10), $\Delta_{n,i}^{1+\delta_1} r(x)$, does not depend on the parameter θ.

The following theorems give a general consistency result and the asymptotic normality of the estimator $\hat{\theta}_n$.

Theorem 1. *(Consistency) Suppose that Assumptions 1 to 7, A_{Step} and Ad hold. Then the estimator $\hat{\theta}_n$ is consistent in probability:*

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0, \qquad n \to \infty.$$

Theorem 2. *(Asymptotic normality) Suppose that Assumptions 1 to 8, A_{Step} and Ad hold. Then*

$$(\sqrt{T_n}(\hat{\mu}_n - \mu_0), \sqrt{n}(\hat{\sigma}_n - \sigma_0)) \xrightarrow{\mathcal{L}} N(0, K) \quad \text{for } n \to \infty,$$

where $K = \begin{pmatrix} (\int_{\mathbb{R}} (\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2 \pi(dx))^{-1} & 0 \\ 0 & 2(\int_{\mathbb{R}} (\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2 \pi(dx))^{-1} \end{pmatrix}$.

The proof of our main results will be presented in Section 8.

It is worth remarking here that, when σ is known and only the parametric estimation of the drift is considered, the model (3) is LAN with Fisher information $I(\mu) = \int_{\mathbb{R}} \frac{(\dot{b}(\theta,x))^2}{a^2(x)} \pi^{\mu}(dx)$ (see [15]).
The Hájek−LeCam convolution theorem states that any regular estimator in a parametric model which satisfies LAN property is asymptotically equivalent to a sum of two independent random variables, one of which is normal with asymptotic variance equal to the inverse of Fisher information, and the other having arbitrary distribution. The efficient estimators are those with the second component identically equal to zero. Therefore, the estimator $\hat{\mu}_n$ is asymptotically efficient in the sense of the Hájek-Le Cam convolution theorem when the volatility parameter σ is supposed to be known and also without its knowledge.
In the general case we are not aware of any result which proves that our model (3) is LAN and therefore nothing can be said about the efficiency of the estimator $\hat{\theta}_n$. However, Gobet has proved in [16] that, in absence of jumps, the LAN property of the model is satisfied with a Fisher information which matches with the variance matrix we found, in presence of jumps, for the joint estimation of the drift and the volatility parameters.
Such a result gives hope for an eventual efficiency of the estimator $\hat{\theta}_n$ here proposed.

4 Practical implementation of the contrast method

In order to use in practice the contrast function (4), one need to know the values of the quantities $m(\mu,\sigma,X_{t_i})$ and $m_2(\mu,\sigma,X_{t_i})$. Even if in most cases, it seems impossible to find an explicit expression for them, explicit or numerical approximations of this functions seem available in many situations.

4.1 Approximation of the contrast function

Let us assume that one has at disposal an approximation of the functions $m(\mu,\sigma,x)$ and $m_2(\mu,\sigma,x)$, denoted by $\tilde{m}(\mu,\sigma,x)$ and $\tilde{m}_2(\mu,\sigma,x)$ which satisfy, for $|x| \leq {\Delta_{n,i}}^{-k_0}$, the following assumptions.
$A\rho$ **:**

1. $|\tilde{m}(\mu,\sigma,x) - m(\mu,\sigma,x)| \leq R(\theta, {\Delta_{n,i}}^{\rho_1}, x), \quad |\tilde{m}_2(\mu,\sigma,x) - m_2(\mu,\sigma,x)| \leq R(\theta, {\Delta_{n,i}}^{\rho_2}, x)$, where the constants $\rho_1 > 1$ and $\rho_2 > 1$ assess the quality of the approximation.

2. $|\frac{\partial^i}{\partial\mu^i}\tilde{m}(\mu,\sigma,x) - \frac{\partial^i}{\partial\mu^i}m(\mu,\sigma,x)| + |\frac{\partial^i}{\partial\sigma^i}\tilde{m}_2(\mu,\sigma,x) - \frac{\partial^i}{\partial\sigma^i}m_2(\mu,\sigma,x)| \leq R(\theta, {\Delta_{n,i}}^{1+\epsilon}, x), \quad$ for $i = 1, 2$, for all $|x| \leq {\Delta_{n,i}}^{-k_0}$ and where $\epsilon > 0$.

3. The bounds on the derivatives of m and m_2 gathered in Propositions 8, 9 and 10 hold true for $\tilde{m}$ and $\tilde{m}_2$ replacing m and m_2.

We have to act on the derivatives of the two approximated functions $\tilde{m}$ and $\tilde{m}_2$ as we do on m and m_2. That is the reason why we need to add the third technical assumption here above, which assure we can move from the derivatives of the real functions to the approximated ones committing an error which is negligible. Now, we consider $\widetilde{\theta}_n$ the estimator obtained from minimization of the contrast function (4) where one has replaced the functions $m(\mu,\sigma,X_{t_i})$ and $m_2(\mu,\sigma,X_{t_i})$ by their approximations $\tilde{m}(\mu,\sigma,x)$ and $\tilde{m}_2(\mu,\sigma,x)$:

$$\tilde{U}_n(\mu,\sigma) := \sum_{i=0}^{n-1}[\frac{(X_{t_{i+1}} - \tilde{m}(\mu,\sigma,X_{t_i}))^2}{\tilde{m}_2(\mu,\sigma,X_{t_i})} + \log(\frac{\tilde{m}_2(\mu,\sigma,X_{t_i})}{\Delta_{n,i}})]\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}},$$

$$\widetilde{\theta}_n = (\tilde{\mu}_n, \tilde{\sigma}_n) \in \arg\min_{(\mu,\sigma)\in\Theta} \tilde{U}_n(\mu,\sigma).$$

Then, the result of Theorem 2 can be extended as follows.

Proposition 1. *Suppose that Assumptions 1 to 8, Ad, A_{Step} and $A\rho$ hold, with $0 < k < k_0$, and that $\sqrt{n}\Delta_n^{\rho_1 - 1/2} \to 0$ and $\sqrt{n}\Delta_n^{\rho_2 - 1} \to 0$ as $n \to \infty$.*
Then, the estimator $\widetilde{\theta}_n := (\tilde{\mu}_n, \tilde{\sigma}_n)$ is asymptotically normal:

$$(\sqrt{T_n}(\tilde{\mu}_n - \mu_0), \sqrt{n}(\tilde{\sigma}_n - \sigma_0)) \xrightarrow{\mathcal{L}} N(0, K) \quad \text{for } n \to \infty,$$

where K is the matrix defined in Theorem 2.

Proposition 1 will be proven in Section 8.4.

We give below several examples of approximations of $m_2(\mu,\sigma,X_{t_i})$ which can be used, together with the approximations of $m(\mu,\sigma,X_{t_i})$ given in Proposition 2 of [2], to construct an explicit contrast function.

4.2 Development of $m_2(\mu,\sigma,x)$.

We provide two kinds of expansion for the function m_2. First, we prove high order expansions that involve only the continuous part of the generator of the process and necessitate the choice of oscillating functions φ. Second, we find an expansion up to order Δ_n^2 for any function φ, and, in particular, show the validity of the condition *Ad* in a general setting. For completeness, we recall also the expansions of the function m found in [2].

4.2.1 Arbitrarily high expansion with oscillating truncation functions.

We show we can write an explicit development for the function m_2, as we did for the function m in Proposition 2 of [2], taking a particular oscillating function φ. In this way, it is therefore possible to make the contrast explicit with approximation at any order. We define $A_{K_1}^{(k)}(x) := \bar{A}_c^k(h_1)(x)$ and

$A^{(k)}_{K_2}(x) := \bar{A}^k_c(h_2)(x)$, where $\bar{A}_c(f) := \bar{b}f' + \frac{1}{2}a^2 f''$, with $\bar{b}(\mu, y) = b(\mu, y) - \int_{\mathbb{R}} \gamma(y) z F(z) dz$; K_1 and K_2 we have written here above stand for "Kessler", based on the fact that the development we find is the same obtained in [24] in the case without jumps by the iteration of the continuous generator $\bar{A}_c$. The functions who appear in the definition of $A^{(k)}_{K_1}$ and $A^{(k)}_{K_2}$ are the following: $h_1(y) := (y - x)$, $h_2(y) = y^2$.
To get Proposition 2 we need to add the following assumption:

Af: We assume that $x \mapsto a(x, \sigma)$, $x \mapsto b(x, \mu)$ and $x \mapsto \gamma(x)$ are $\mathcal{C}^\infty$ functions, they have at most uniform in μ and σ polynomial growth as well as their derivatives.

Proposition 2. *Assume that Assumptions 1-4 and Af hold and let φ be a $\mathcal{C}^\infty$ function that has compact support and such that $\varphi \equiv 1$ on $[-1, 1]$ and $\forall k \in \{0, ..., M\}$, $\int_{\mathbb{R}} x^k \varphi(x) dx = 0$ for $M \geq 0$. Moreover we suppose that the Lévy density F is $\mathcal{C}^\infty$. Then, for $|x| \leq \Delta_{n,i}^{-k_0}$ with some $k_0 > 0$,*

$$m_2(\mu, \sigma, x) = \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A^{(k)}_{K_2}(x) \frac{\Delta^k_{n,i}}{k!} - (x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A^{(k)}_{K_1}(x) \frac{\Delta^k_{n,i}}{k!})^2 + R(\theta, \Delta^{\beta(M+2)}_{n,i}, x). \tag{11}$$

Moreover, for $\vartheta = \mu$ or $\vartheta = \sigma$

$$|\frac{\partial}{\partial \vartheta} R(\theta, \Delta^{\beta(M+2)}_{n,i}, x)| \leq R(\theta, \Delta^{\beta(M+2)}_{n,i}, x). \tag{12}$$

It is proved below Proposition 2 in [2] that a function φ which satisfies the assumptions here above exists: it is possible to build it through ψ, a function with compact support, $\mathcal{C}^\infty$ and such that $\psi|_{[-1,1]}(x) = \frac{x^M}{M!}$. It is enough to define $\varphi(x) := \frac{\partial^M}{\partial x^M}\psi(x)$ to get $\varphi \equiv 1$ on $[-1, 1]$; φ is $\mathcal{C}^\infty$, with compact support and such that for each $l \in \{0, ...M\}$, using the integration by parts, $\int_{\mathbb{R}} x^l \varphi(x) dx = 0$.
It is thanks to such a choice of an oscillating function φ that the contribution of the discontinuous part of the generator disappears and we get the same development found in the continuous case, in Kessler [24], due only to the continuous generator.
In the situation where (11) holds true with $\lfloor \beta(M + 2) \rfloor > 2$, we get a development for m_2 as in *Ad* for $r(x)$ identically 0 and $r(\mu, \sigma, x)$ being an explicit function.

For completeness, let us recall that under the same assumptions as in Proposition 2, we have the following expansion for m (see Proposition 2 in [2]):

$$m(\mu, \sigma, x) = x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A^{(k)}_{K_1}(x) \frac{\Delta^k_{n,i}}{k!} + R(\theta, \Delta^{\beta(M+2)}_{n,i}, x), \text{ for } |x| \leq \Delta^{-k_0}_{n,i}, \text{ with } k_0 > 0. \tag{13}$$

In the definition of the contrast function we can replace the two functions m and m_2 with their approximations obtained from (11) and (13) :

$$\tilde{m}^{\rho_1}(\mu, \sigma, x) := x + \sum_{k=1}^{\rho_1} A^{(k)}_{K_1}(x) \frac{\Delta^k_{n,i}}{k!} \quad \text{and}$$

$$\tilde{m_2}^{\rho_2}(\mu, \sigma, x) := \sum_{k=1}^{\rho_2} A^{(k)}_{K_2}(x) \frac{\Delta^k_{n,i}}{k!} - (x + \sum_{k=1}^{\rho_2} A^{(k)}_{K_1}(x) \frac{\Delta^k_{n,i}}{k!})^2.$$

The errors we commit are, respectively, $R(\theta, \Delta^{\rho_1}_{n,i}, x)$ and $R(\theta, \Delta^{\rho_2}_{n,i}, x)$; for $\rho_i \leq \lfloor \beta(M + 2) \rfloor$ and $\rho_i \geq 1$, $i = 1, 2$.
By application of the Proposition 1 we can see that the associated estimator is asymptotically gaussian under the assumptions $\sqrt{n}\Delta_n^{\rho_1 - \frac{1}{2}} \to 0$ and $\sqrt{n}\Delta_n^{\rho_2 - 1} \to 0$, for n going to ∞. Since M and thus ρ_1 and ρ_2 can be chosen arbitrarily large, we see that the sampling step Δ_n is allowed to go to zero in a arbitrary slowly polynomial rate as function of n. That means that a slow sampling step needs to choose a truncation function with more vanishing moments.

4.2.2 Second order expansion with general truncation functions.

Another situation in which *Ad* holds is gathered in the following proposition, that will be still proven in the appendix:

Proposition 3. *Suppose that Assumptions A1 -A5 and A7 hold, that $\beta \in (\frac{1}{4}, \frac{1}{2})$ and that the Lévy density F is $\mathcal{C}^1$. Then there exists $k_0 > 0$ such that, for $|x| \leq \Delta^{-k_0}_{n,i}$,*

$$m_2(\mu, \sigma, x) = \Delta_{n,i} a^2(x, \sigma) + \frac{\Delta^{1+3\beta}_{n,i}}{\gamma(x)} F(0) \int_{\mathbb{R}} v^2 \varphi(v) dv + \Delta^2_{n,i}(3\bar{b}^2(x, \mu) + h_2(x, \theta)) + \Delta^{(1+4\beta)\wedge(2+\beta)\wedge(3-2\beta)}_{n,i} R(\theta, 1, x); \tag{14}$$

where $h_2 = \frac{1}{2}a^2(a')^2 + \frac{1}{2}a^3a'' + a^2\bar{b}' + aa'\bar{b} + \bar{b}^2$. *Moreover, for both* $\vartheta = \mu$ *and* $\vartheta = \sigma$, $\frac{\partial}{\partial\vartheta}R(\theta, 1, x)$ *is still a* $R(\theta, 1, x)$ *function.*

We observe that, defining $r(x) := F(0)\int_{\mathbb{R}} v^2\varphi(v)dv$ and $r(\mu, \sigma, x) := \frac{3\bar{b}^2(x,\mu)+h_2(x,\theta)}{a^2(x,\sigma)}$, we get a development as in *Ad* for $\delta_1 := 3\beta$.
We observe that if $\int_{\mathbb{R}} v^2\varphi(v)dv = 0$, we fall back in development of Proposition 2 up to order 2. We therefore see that the choice of an oscillating truncated function φ is necessary in order to remove the jump contribution.

It is worth noting here that biggest term after the main one is due to the jump part and do not depend on the parameters μ and σ. We will see in the sequel that is necessary, in order to prove the consistency of $\hat{\mu}_n$, that this contribution does not depend on the drift parameter. Considering indeed the difference of the contrast computed for two different values of the drift parameter, its presence results irrelevant.

We remark that the term with order $1 + 4\beta$ is negligible compared to the order 2 terms since in our setting β is assumed to be bigger than $\frac{1}{4}$.

In Proposition 3 before F is required to be $\mathcal{C}^1$, such assumption is no longer needed in following more general proposition.

Proposition 4. *Suppose that Assumptions A1 -A5 and A7 hold. Then there exists* $k_0 > 0$ *such that, for* $|x| \leq \Delta_{n,i}^{-k_0}$,

$$m_2(\mu, \sigma, x) = \Delta_{n,i}a^2(x, \sigma) + \frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)}\int_{\mathbb{R}} u^2\varphi(u)F(u\frac{\Delta_{n,i}^{\beta}}{\gamma(x)})du + \Delta_{n,i}^2(3\bar{b}^2(x, \mu) + h_2(x, \theta)) +$$

$$+\frac{\Delta_{n,i}^{2+\beta}a^2(x, \sigma)}{2\gamma(x)}\int_{\mathbb{R}}(u\varphi'(u) + u^2\varphi''(u))F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du + \Delta_{n,i}^{(3-2\beta)\wedge(2+\beta)}R(\theta, 1, x), \tag{15}$$

where $h_2 = \frac{1}{2}a^2(a')^2 + \frac{1}{2}a^3a'' + a^2\bar{b}' + aa'\bar{b} + \bar{b}^2$.
Moreover, for both $\vartheta = \mu$ *and* $\vartheta = \sigma$, $\frac{\partial}{\partial\vartheta}R(\theta, 1, x)$ *is still a* $R(\theta, 1, x)$ *function.*

We see that the contributions of the jumps depend on the density F which argument in the integral depend on $\Delta_{n,i}$. If we choose a particular density function F which is null in the neighborhood of 0 the contribution of the jumps disappears and, in this case, we fall on the development for m_2 found by Kessler in the case without jumps ([24]), up to order $\Delta_{n,i}^2$.

The expansion (15) looks cumbersome, however all terms are necessary to get an expansion with a remainder term of explicit order strictly greater than 2, and valid for any finite intensity F. In the particular case where F is $\mathcal{C}^1$, the first three terms in the expansion give the the main terms of the expansions (14), while the last integral term, with order $\Delta_{n,i}^{2+\beta}$, is clearly a rest term. However, in the situation where F may be unbounded near 0, with $\int F(z)dz < \infty$, the last integral term is only seen to be negligible versus $\Delta_{n,i}^2$. Hence, this last integral term may be non negligible compared to the rest term and is needed in the expansion.

We remark that the construction of the contrast function requires the knowledge of the law of the jumps, which is typically unknown, since it affects m and m_2. However, Propositions 2 and 3 suggest us some suitable approximation of m ad m_2 for which the corrections become explicit. For example, requiring also that $\int_{\mathbb{R}} zF(z) = 0$, $\bar{b}$ turns out being simply b and therefore $\bar{A}_c(f)$ no longer depends on the jump law. It implies that A_{k_1} in Proposition 2 (and thus the expansions (11) and (13)) no longer depend on the jump law either. Moreover, Proposition 3 suggests us to estimate in a non -parametric way the quantity $F(0)$, which would provide an explicit correction as well. Even if we are not aware of any method for the estimation of $F(0)$, there are many papers related to such a topic. The problem of the estimation of $F(x)$ with $x \neq 0$ is connected for example to the works of [8] in high frequency and of [17] in low frequency, considering in both cases a Lévy process. In [23] the authors deal instead with a Lévy driven Ornstein Uhlenbeck (OU) process. Figueroa - López proposes nonparametric estimators for the Lévy density and obtains in [12] both pointwise and uniform central limit theorems on an interval away from the origin for these estimators.

Finally, we recall the expansion of m valid under the Assumptions A1-A4 (see Theorem 2 in [2]),

$$m(\mu, \sigma, x) = x + \Delta_{n,i}\bar{b}(x, \mu) + \frac{\Delta_{n,i}^{1+2\beta}}{\gamma(x)}\int_{\mathbb{R}} u\varphi(u)F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du + R(\theta, \Delta_{n,i}^{2-2\beta}, x). \tag{16}$$

Developments of m_2 given in Propositions 2 - 4, together with the developments of m given in (13), (16) will be useful for the applications, as illustrated in the following section.

5 Simulation study

In this section we provide some numerical results. First of all, in Section 5.1, we consider a second order expansion for m and m_2 which holds true for any truncation function φ. After that, in Section 5.2, some oscillating truncation functions will be employed with the purpose of using an arbitrarily high expansion for m and m_2, as in Kessler. As we will see, it will lead us to some estimators very well-performed.

5.1 Second order expansion

Let us consider the model

$$X_t = X_0 + \int_0^t (\theta_1 X_s + \theta_2)ds + \sigma W_t + \gamma \int_0^t \int_{\mathbb{R}\setminus\{0\}} z\tilde{\mu}(ds, dz), \tag{17}$$

where the compensator of the jump measure is $\overline{\mu}(ds, dz) = \lambda F_0(z)dsdz$ for F_0 the probability density of the law $\mathcal{N}(\mu_J, \sigma_J^2)$ with $\mu_J \in \mathbb{R}$, $\sigma_J > 0$, $\sigma > 0$, $\theta_1 < 0$, $\theta_2 \in \mathbb{R}$, $\gamma \geq 0$, $\lambda \geq 0$.
We want to explore approximations for m and m_2 which make us able to find en explicit version of the contrast function. According to Theorem 2 of [2] (see also (16)) we have, considering a threshold level which is $c\Delta_{n,i}^{\beta}$, the following development for m:

$$m(\theta_1, \theta_2, \sigma, x) = x + \Delta_{n,i}\bar{b}(x, \theta_1, \theta_2) + \frac{c^2\Delta_{n,i}^{1+2\beta}}{\gamma}\int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_{n,i}^{\beta}}{\gamma})du + R(\theta, \Delta_{n,i}^{2-2\beta}, x);$$

which leads us to the approximation

$$\widetilde{m}(\theta_1, \theta_2, x) = x + \Delta_{n,i}\bar{b}(x, \theta_1, \theta_2) + \frac{c^2\Delta_{n,i}^{1+2\beta}}{\gamma}\int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_{n,i}^{\beta}}{\gamma})du,$$

for $\bar{b}(x, \theta_1, \theta_2) = (\theta_1 x + \theta_2) - \gamma\lambda\mu_j$. The sampling scheme is uniform and therefore $\Delta_{n,i} = \frac{T}{n}$, for each $i \in \{0, ...n-1\}$. It follows $|m(\theta_1, \theta_2, \sigma, x) - \widetilde{m}(\theta_1, \theta_2, x)| \leq R(\theta, \Delta_{n,i}^{2-2\beta}, x) = R(\theta, \Delta_{n,i}^{\rho_1}, x)$.
Concerning the approximation of m_2, from its development gathered in Proposition 4 we define

$$\widetilde{m}_2(\sigma) := \Delta_{n,i}\sigma^2 + \frac{\Delta_{n,i}^{1+3\beta}c^3}{\gamma}\int_{\mathbb{R}} u^2\varphi(u)F(\frac{uc\Delta_{n,i}^{\beta}}{\gamma})du,$$

which is such that $|m_2(\theta_1, \theta_2, \sigma, x) - \widetilde{m}_2(\sigma)| \leq R(\theta, \Delta_{n,i}^2, x) = R(\theta, \Delta_{n,i}^{\rho_2}, x)$.
By application of Proposition 1 we see that the associated estimator is asymptotically gaussian under the strong assumption $\sqrt{n}\Delta_n^{\frac{3}{2}-2\beta} \to 0$, for $n \to \infty$, with $\beta \in (\frac{1}{4}, \frac{1}{2})$. In the best case β is close to $\frac{1}{4}$ and so the condition here above becomes equivalent to $n\Delta_n^2 \to 0$. Despite such a strong requirement on the discretization step, we expect the proposed method performing better than the Euler schema since the main contribution of the jumps has there already been removed.

We want to compare the estimator $\tilde{\theta}_n$ we get by the minimization of the contrast function obtained by the expansions of m and m_2 provided here above with the estimator based on the Euler scheme approximation:

$$\widetilde{m}^E(\theta_1, \theta_2, x) = x + (\theta_1 x + \theta_2 - \lambda\gamma\mu_j)\Delta_{n,i}, \qquad \widetilde{m}_2^E(\sigma, x) = \sigma^2\Delta_{n,i}. \tag{18}$$

We estimate jointly the parameter $\theta = (\theta_1, \theta_2, \sigma)$ by minimization of the contrast function

$$U_n(\theta) = \sum_{i=0}^{n-1}[\frac{(X_{t_{i+1}} - \widetilde{m}(\theta_1, \theta_2, X_{t_i}))^2}{\widetilde{m}_2(\sigma, X_{t_i})} + \log(\frac{\widetilde{m}_2(\theta_1, \sigma, X_{t_i})}{\Delta_{n,i}})]\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}), \tag{19}$$

where $c > 0$ will be specified later.
We compute the derivatives of the contrast function with respect to the three parameters:

$$\frac{\partial}{\partial\theta_1}U_n(\theta) = \sum_{i=0}^{n-1}\frac{2(X_{t_{i+1}} - \widetilde{m}(\theta_1, \theta_2, X_{t_i}))X_{t_i}}{\widetilde{m}_2(\sigma)}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) =$$

$$= \frac{2}{\widetilde{m}_2(\sigma)}\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i} - \Delta_{n,i}\theta_1 X_{t_i} - \Delta_{n,i}\theta_2 + \Delta_{n,i}\gamma\lambda\mu_j - J_1^i)X_{t_i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}),$$

where we have denoted as J_1^i the term in the development of $\widetilde{m}$ turning up from the presence of jumps, which is $\frac{c^2\Delta_{n,i}^{1+2\beta}}{\gamma}\int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_{n,i}^{\beta}}{\gamma})du$.
We want $\frac{\partial}{\partial\theta_1}U_n(\theta)=0$, it leads us to the definition of the following estimator:

$$\tilde{\theta}_{1,n} := \frac{\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i} - \Delta_{n,i}\theta_2 + \Delta_{n,i}\gamma\lambda\mu_j - J_1^i)X_{t_i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}X_{t_i}^2\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}.$$

In the same way

$$\frac{\partial}{\partial\theta_2}U_n(\theta) = \sum_{i=0}^{n-1}\frac{2(X_{t_{i+1}} - \widetilde{m}(\theta_1,\theta_2,X_{t_i}))}{\tilde{m}_2(\sigma)}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) =$$

$$= \frac{2}{\widetilde{m}_2(\sigma)}\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i} - \Delta_{n,i}\theta_1 X_{t_i} - \Delta_{n,i}\theta_2 + \Delta_{n,i}\gamma\lambda\mu_j - J_1^i)\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}).$$

Since we want $\frac{\partial}{\partial\theta_2}U_n(\theta)=0$, we define $\tilde{\theta}_{2,n}$ in the following way:

$$\tilde{\theta}_{2,n} := \frac{\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i} - \Delta_{n,i}\theta_1 X_{t_i})\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})} + \gamma\lambda\mu_j - \frac{\sum_{i=0}^{n-1}J_1^i\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})} =$$

$$= \tilde{\theta}_n^{\text{Euler}} - \frac{\sum_{i=0}^{n-1}J_1^i\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}; \tag{20}$$

we can see $\tilde{\theta}_{2,n}$ as a corrected version of the estimator $\tilde{\theta}_n^{\text{Euler}}$ that would result considering the Euler scheme approximation for the function m, as in (18). We observe moreover that, considering a uniform discretization step, the last term in (20) becomes simply $\frac{J_1}{\Delta_n} := \frac{c^2\Delta_n^{2\beta}}{\gamma}\int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_n^{\beta}}{\gamma})du$.
Computing also the derivative of the contrast function with respect to σ^2, we have

$$\frac{\partial}{\partial\sigma^2}U_n(\theta) = \sum_{i=0}^{n-1}[\frac{-(X_{t_{i+1}} - \widetilde{m}(\theta_1,\theta_2,X_{t_i}))^2\partial_{\sigma^2}\widetilde{m}_2(\sigma) + \widetilde{m}_2(\sigma)\partial_{\sigma^2}\widetilde{m}_2(\sigma)}{\widetilde{m}_2^2(\sigma)}]\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}),$$

which is equal to zero if and only if

$$\sum_{i=0}^{n-1}[-(X_{t_{i+1}} - \widetilde{m}(\theta_1,\theta_2,X_{t_i}))^2\Delta_{n,i} + \Delta_{n,i}(\Delta_{n,i}\sigma^2 + J_2^i)]\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) = 0,$$

for $J_2^i := \frac{\Delta_{n,i}^{1+3\beta}c^3}{\gamma}\int_{\mathbb{R}} u^2\varphi(u)F(\frac{uc\Delta_{n,i}^{\beta}}{\gamma})du$, which is the part deriving from the jumps in the development of $\widetilde{m}_2(\sigma)$. It drives us to the estimator

$$\tilde{\sigma}_n^2 := \frac{\sum_{i=0}^{n-1}(X_{t_{i+1}} - \tilde{m}(\theta_1,\theta_2,X_{t_i}))^2\Delta_{n,i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}^2\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})} - \frac{\sum_{i=0}^{n-1}\Delta_{n,i}J_2^i\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}^2\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}.$$

Considering an uniform discretization step it is

$$\tilde{\sigma}_n^2 := \frac{\sum_{i=0}^{n-1}(X_{t_{i+1}} - \tilde{m}(\theta_1,\theta_2,X_{t_i}))^2\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\Delta_n\sum_{i=0}^{n-1}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})} - \frac{J_2}{\Delta_n}.$$

Again, it can be seen as a corrected version of $\tilde{\sigma}_n^{2,\text{Euler}}$, the estimator that would have resulted considering the approximation of the functions m and m_2 as defined in (18). In such a case, not only we would not have seen the contribution of the jumps appearing in the second term here above, but also in the first term we should have replaced $\tilde{m}$ with its Euler approximation.
To illustrate the estimation method, we focus on the estimation of the parameters θ_2 and σ^2 only.
For numerical simulations we choose $T = 100$, $n = 50\,000$, $\lambda = 1$, $\gamma = 1$, $\theta_1 = -1$, $\theta_2 = 2$, $\sigma = 0.5$, $d = 2$ and $c = 1.25$. We estimate the bias of the estimators using a Monte Carlo method based on 1000 replications.
First, we consider β as big as possible, fixing it equal to 0.49; then we will take $\beta = 0.3$; in both cases the jumps size has common law $\mathcal{N}(4, 0.25)$.

	Mean (dev std) for $\theta_2 = 2$	Mean (dev std) for $\sigma^2 = 0.25$
$\tilde{\theta}_n^{\text{Euler}}$	2.00284 (0.05020)	0.26558 (0.00161)
$\widetilde{\theta}_n$	2.00289 (0.04830)	0.26412 (0.00156)

Table 1: Monte Carlo estimates of θ_2 and σ^2 from 1000 samples. We have here fixed $\beta = 0.3$.

	Mean (dev std) for $\theta_2 = 2$	Mean (dev std) for $\sigma^2 = 0.25$
$\tilde{\theta}_n^{\text{Euler}}$	2.00368 (0.05195)	0.25150 (0.00161)
$\widetilde{\theta}_n$	2.00302 (0.04992)	0.25067 (0.00159)

Table 2: Monte Carlo estimates of θ_2 and σ^2 from 1 000 samples. We have here fixed $\beta = 0.49$.

We see that there is not a big difference in the estimation of θ_2 with and without the correction term. The two estimators of θ_2 differs in fact only for the contribution of the jumps $\frac{J_1}{\Delta_n} = \frac{c^2\Delta_n^{2\beta}}{\gamma}\int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_n^{\beta}}{\gamma})du$ that is in this case close to zero, because of the natural choice of taking a truncated function which is symmetric. Indeed, even if the density function F is not symmetric, asymptotically the only contribution it gives is due by its value in zero and, therefore, the symmetry of φ is enough to ensure the limit of the integral is zero.
Regarding σ^2, as the correction term $\frac{J_2}{\Delta_n}$ is very small the two different estimators provide two means for σ^2 which are rather similar. However, the estimator we find through our approximated functions $\tilde{m}$ and $\tilde{m}_2$ performs a bit better than the estimator we got through Euler scheme.

5.2 Kessler approximation

In the previous case we have used the second order expansion from m and m_2, available for any truncation function φ. Now we still consider (17) in which F_0 is still the probability density of the law $\mathcal{N}(\mu_j, \sigma_J^2)$, but we use Kessler approximation to remove the bias. To do that, we take an oscillating truncated function φ for which the initial contributions of the discontinuous part of the generator disappear.
Indeed, according to Proposition 2 (respectively Proposition 2 in [2]) we know that, using sufficiently oscillating truncation functions, the expansion of m_2 (respectively m) is the same found by Kessler in the continuous case. The usefulness of Proposition 2 is to illustrate it is possible to neglect the effect of the truncation function φ as the expansions of m and m_2 are identical to those of Kessler for the continuous part of the SDE. Since the Kessler's expansion approximates the first conditional moments of $\bar{X}_t = \bar{X}_0 + \int_0^t(\theta_1\bar{X}_s + \theta_2 - \gamma\lambda\mu_J)ds + \sigma W_t$ (see (1)), which is the continuous part of (17) and which is explicit due to the linearity of the model, we decide to use directly the expression of the conditional moment and set

$$\widetilde{m}(\theta_1, \theta_2, x) = (x + \frac{\theta_2}{\theta_1} - \frac{\gamma\lambda\mu_J}{\theta_1})e^{\theta_1\Delta_{n,i}} + \frac{\gamma\lambda\mu_J - \theta_2}{\theta_1}; \tag{21}$$

while the approximation of $m_2(\mu, \sigma, x)$ is

$$\widetilde{m}_2(\theta_1, \sigma, x) = \frac{\sigma^2}{2\theta_1}(e^{2\theta_1\Delta_{n,i}} - 1). \tag{22}$$

We want to compare the estimator $\tilde{\theta}_n$ we get by the minimization of the contrast function obtained by the Kessler exact correction of the bias in which we use the approximations (21) and (22) for m and m_2 with the estimator based on the Euler scheme approximation. Following Nikolskii [31], we construct oscillating truncation functions in the following way. First, we choose $\varphi^{(0)} : \mathbb{R} \to [0, 1]$ a $\mathcal{C}^\infty$ symmetric function with support on $[-2, 2]$ such that $\varphi^{(0)}(x) = 1$ for $|x| \leq 1$. We let, for $d > 1$, $\varphi_d^{(1)}(x) = (d\varphi^{(0)}(x) - \varphi^{(0)}(x/d))/(d-1)$, which is a function equal to 1 on $[-1, 1]$, vanishing on $[-d, d]^c$ and such that $\int_{\mathbb{R}} \varphi_d^{(1)}(x)dx = 0$. For $l \in \mathbb{N}$, $l \geq 1$, and $d > 1$, we set $\varphi_d^{(l)}(x) = c_d^{-1}\sum_{k=1}^l C_l^k(-1)^{k+1}\frac{1}{k}\varphi_d^{(1)}(x/k)$, where $c_d = \sum_{k=1}^l C_l^k(-1)^{k+1}\frac{1}{k}$. One can check that $\varphi_d^{(l)}$ is compactly supported, equal to 1 on $[-1, 1]$, and that for all $k \in \{0, \ldots, l\}$, $\int_{\mathbb{R}} x^k\varphi_d^{(l)}(x)dx = 0$, for $l \geq 1$. With these notations, we estimate the parameter $(\theta_1, \theta_2, \sigma)$ by minimization of the contrast function (19), implementing the just built truncation function $\varphi^{(l)}_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})$, where $l \in \mathbb{N}$ and $c > 0$ will be specified latter.

For numerical simulations, we choose $T = 1\,000$, n will be chosen equal to 2 000, 10 000 and 50 000, $\theta_1 = -1$, $\theta_2 = 2$, $\sigma = 0.5$ and $X_0 = x_0 = 0$. We estimate the bias and standard deviation of our estimators using a Monte Carlo method based on 2 000 replications. As a start, we consider a situation

without jumps $\lambda = 0$, in which we remove the truncation function φ in the contrast, as it is useless in absence of jumps.

	Mean (std) for $\theta_1 = -1$	Mean (std) for $\theta_2 = 2$	Mean (std) for $\sigma = 0.5$
$\Delta_n = 0.5$			
$\tilde{\theta}_n^{\text{Euler}}$	-0.7922 (0.0348)	1.5843 (0.0701)	0.3980 (0.0062)
$\widetilde{\theta}_n$	-1.0063 (0.0578)	2.0128 (0.1166)	0.5017 (0.0099)
$\Delta_n = 0.1$			
$\tilde{\theta}_n^{\text{Euler}}$	-0.9536 (0.0422)	1.9068 (0.0857)	0.4761 (0.0034)
$\widetilde{\theta}_n$	-1.0053 (0.0465)	2.0106 (0.0944)	0.5003 (0.0037)
$\Delta_n = 0.02$			
$\tilde{\theta}_n^{\text{Euler}}$	-0.9940 (0.0442)	1.9888 (0.0895)	0.4951 (0.0016)
$\widetilde{\theta}_n$	-1.0039 (0.0437)	2.0084 (0.0886)	0.5001 (0.0016)

Table 3: Process without jump

In Table 3, we compare the estimator which uses the Kessler exact bias corrections given by (21) and (22), with an estimator based on the Euler scheme approximation. From Table 3 we see that the estimator $\tilde{\theta}_n^{\text{Euler}}$ based on Euler contrast exhibits some bias which is completely removed using Kessler's correction. The improvement provided by Kessler approximation is particularly evident for n small.

Next, we set a jump intensity $\lambda = 0.1$, with jumps size whose common law is $\mathcal{N}(4, 0.25)$ and set $\gamma = 1$. Also in this case, we compare the results obtained basing the estimation on Kessler or on Euler approximations for the quantities m and m_2. Numerical simulations are provided for the truncation function $\varphi^{(0)}$, where $c = 2$ and $\beta = 0,49$. A plot of this function can be found in Figure 1.

	Mean (std) for $\theta_1 = -1$	Mean (std) for $\theta_2 = 2$	Mean (std) for $\sigma = 0.5$
$\Delta_n = 0.5$			
$\tilde{\theta}_n^{\text{Euler}}$	-0.7301 (0.0274)	1.5796 (0.0512)	0.4843 (0.0223)
$\widetilde{\theta}_n$	-0.9095 (0.0423)	1.8697 (0.0763)	0.5967 (0.0234)
$\Delta_n = 0.1$			
$\tilde{\theta}_n^{\text{Euler}}$	-0.9509 (0.0153)	1.9213 (0.0338)	0.4761 (0.0034)
$\widetilde{\theta}_n$	-0.9988 (0.0178)	1.9974 (0.0390)	0.5002 (0.0035)
$\Delta_n = 0.02$			
$\tilde{\theta}_n^{\text{Euler}}$	-0.9909 (0.0165)	1.9858 (0.0361)	0.4951 (0.0016)
$\widetilde{\theta}_n$	-1.0013 (0.0166)	2.0028 (0.0374)	0.5000 (0.0016)

Table 4: Gaussian jumps with $\lambda = 0.1$

Results in Table 4 show that the estimator deriving from Kessler approximation works well, better than the one deriving from Euler approximations. The bias is visibly reduced for all the choices of n, especially when n is small. It matches with our theoretical results for which, implementing an oscillating truncation function, the discretization step can goes to zero arbitrarily slow. We remark that by the choice of a symmetric truncation function one has $\int_{\mathbb{R}} u\varphi^{(0)}(u)du = 0$ and it can be seen that this conditions is sufficient, in the expansion of m and m_2, to suppress the largest contribution of the discrete part of the generator.

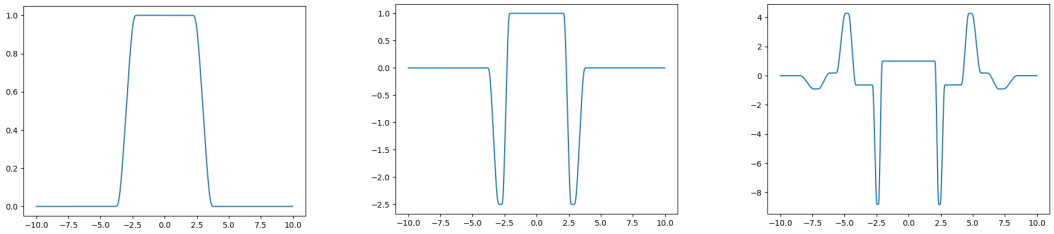
When the number of jumps is greater, e.g. for $\lambda = 10$, we choose once again $c = 2$ and $\beta = 0,49$ but, unlike before, T is fixed equal to 100 and n is 10 000, 50 000 and 500 000. Now the law of the jumps is $\mathcal{N}(0, 1)$. We compare the results we get by considering different oscillating truncation functions $\varphi^{(0)}$, $\varphi_{1.4}^{(2)}$ and $\varphi_{1.2}^{(4)}$, whose plots are given in Figure 1.

We see in Table 5 that the use of the more oscillating kernels $\varphi_{1.4}^{(2)}$, $\varphi_{1.2}^{(4)}$ yields to a smaller bias than using $\varphi^{(0)}$.

The estimator we get using $\varphi_{1.2}^{(4)}$ performs particularly well in this situation, it has a negligible bias and a small standard deviation.

	Mean (std) for $\theta_1 = -1$	Mean (std) for $\theta_2 = 2$	Mean (std) for $\sigma = 0.5$
$\Delta_n = 0.01$			
$\widetilde{\theta_n}$ using $\varphi^{(0)}$	-0.9611 (0.0477)	1.9229 (0.1440)	1.1037 (0.0229)
$\widetilde{\theta_n}$ using $\varphi^{(2)}_{1.4}$	-0.9979 (0.0387)	1.9967 (0.1157)	0.3272 (0.0310)
$\widetilde{\theta_n}$ using $\varphi^{(4)}_{1.2}$	-0.9979 (0.1020)	1.9924 (0.3123)	0.5376 (0.1013)
$\Delta_n = 0.002$			
$\widetilde{\theta_n}$ using $\varphi^{(0)}$	-0.9948 (0.0307)	1.9890 (0.0926)	0.7121 (0.0110)
$\widetilde{\theta_n}$ using $\varphi^{(2)}_{1.4}$	-1.0007 (0.0241)	2.0007 (0.0707)	0.4863 (0.0031)
$\widetilde{\theta_n}$ using $\varphi^{(4)}_{1.2}$	-0.9996 (0.0374)	1.9994 (0.1116)	0.5007 (0.0144)
$\Delta_n = 0.0002$			
$\widetilde{\theta_n}$ using $\varphi^{(0)}$	-1.0008 (0.0237)	1.9991 (0.0717)	0.5331 (0.0027)
$\widetilde{\theta_n}$ using $\varphi^{(2)}_{1.4}$	-1.0009 (0.0217)	2.0021 (0.0643)	0.4995 (0.0005)
$\widetilde{\theta_n}$ using $\varphi^{(4)}_{1.2}$	-0.9996 (0.0226)	1.9978 (0.0677)	0.4999 (0.0010)

Table 5: Gaussian jumps with $\lambda = 10$



(a) $\varphi^{(0)}$ (b) $\varphi^{(2)}_{1.4}$ (c) $\varphi^{(4)}_{1.2}$

Figure 1: Plot of the truncation functions

6 Future perspectives

In future perspectives, we plan to generalize the obtained results. There are many remaining questions of interest, both from a theoretical and a numerical point of view.

As explained in the introduction, in this paper we extend previous works in which only the drift parameter was considered (see for example [15] and [2]). We may wonder what happens if the jump coefficient has another parameter, that we here denote as η. A first, simple answer is that, if the jumps are centered (which means $\int_{\mathbb{R}} zF(z)dz = 0$), then the knowledge of the jumps is no longer needed. Hence, as discussed below Proposition 4, it is enough to use the approximation of m and m_2 proposed in Proposition 2 of [2] and in Proposition 2 in order to make the contrast explicit. After that, we can still use our main results to estimate jointly the drift and the volatility parameters and the lack of knowledge of the jump parameter is not a problem. Estimating the three parameters jointly, that is a different story. In [36], Shimizu and Yoshida deal with it. They propose a contrast function which has two terms: the first corresponds to the contrast for an usual diffusion process, while the second concerns the discretization of the likelihood function of a compound Poisson process with Lévy density. In the abovementioned work they show the asymptotic normality of the estimators under some conditions on the sampling step which are more and more restrictive as the intensity of the jumps is high and, when the intensity is finite, they reduces to $n\Delta_n^2 \to 0$.

One may wonder how to modify the contrast function introduced in (4) in order to estimate jointly the three parameters under the general balance condition $n\Delta_n \to \infty$ and $n\Delta_n^k \to 0$ for all $k \geq 2$. The natural idea is to add to (4) a term, corresponding to the one introduced by Shimizu and Yoshida, which regards the discretization of the likelihood function of the jumps. In order to weaken the condition on the sampling step we need to introduce in such a term a third unknown quantity that we would define on the specif purpose to make $\partial_\eta U_n$ a triangular array of martingale increments. However, it is not a simple issue. First of all, it is not trivial to understand how to correct the contrast function and, furthermore, it would become much more difficult to investigate the asymptotic behaviour of the new contrast function and of its derivatives.

For practical implementation, the question of approximation of m and m_2 is crucial, and therefore

one has to face the issue of choosing the threshold level, characterized here by c, β and φ. Indeed, though the efficiency of the estimators has been established theoretically, it is known that in general their real performance strongly depends on a choice of tuning parameters; see for example Shimizu [35], Iacus and Yoshida [19]. The filter is each time based on only one increment of the data and so, in this sense, this filter can be regarded as a local method. In [20], Inatsugu and Yoshida introduce a global filtering method, which they call α threshold method. It uses all of the data to more accurately detect increments having jumps, based on the order statistics associated with all increments. Even if the α - threshold method involves the tuning parameter α to determine a selection rule for increments, it is robust against the choice of such a parameter.
In this work we use a local threshold method, having in mind a different objective. Here the idea, indeed, is mainly to remove the bias caused by a wrong choice of the threshold parameters rather than determine a good one. On contrary to more conventional threshold methods, the quantities m and m_2 depend here by construction on the threshold level and may compensate for too large threshold. However, as seen in Section 5, the quantities m and m_2 can be numerically very far from the approximations derived through the Euler scheme approximation and through the approximation scheme at higher order (as proposed in Proposition 3). A perspective would be to approximate numerically these two quantities, using for instance a Monte Carlo approach, and provide more accurate corrections than the ones here proposed.

7 Preliminary results

Before proving the main statistical results of Section 3, we need to state several propositions which will be useful in the sequel. They will be proven in Section A.1.

7.1 Limit theorems

The asymptotic properties of estimators are deduced from the asymptotic behavior of the contrast function. We therefore state some propositions useful to get the asymptotic behavior of U_n.

Proposition 5. *Suppose that Assumptions 1 to 4 and A_{Step} hold, $\Delta_n \to 0$ and $T_n \to \infty$ and f is a differentiable function $\mathbb{R}\times\Theta \to \mathbb{R}$ such that $|f(x,\theta)| \le c(1+|x|^c)$, $|\partial_x f(x,\theta)| \le c(1+|x|^c)$ and, for $\vartheta = \mu$ and $\vartheta = \sigma$, $|\partial_\vartheta f(x,\theta)| \le c(1+|x|^c)$. Then $x \mapsto f(x,\theta)$ is a π-integrable function for any $\theta \in \Theta$ and the following convergences hold as $n \to \infty$:*
1. $|\frac{1}{T_n}\sum_{i=0}^{n-1} \Delta_{n,i} f(X_{t_i},\theta) 1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0$,
2. $|\frac{1}{T_n}\sum_{i=0}^{n-1} \Delta_{n,i} f(X_{t_i},\theta) \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i}) 1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0$,
3. $|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta) 1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0$,
4. $|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta) \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i}) 1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0$.

Statements $1-2$ and $3-4$ of the proposition here above, as well as the first and the second point of Proposition 6 below, turn out being similar if the sampling step $\Delta_{n,i} = \Delta_n$ considered is uniform. Otherwise, we need these two different convergences because, in order to estimate μ and σ jointly, we have to deal with different scaling of the contrast function.

Proposition 6. *Suppose that Assumptions 1 to 4 and A_{Step} hold, $\Delta_n \to 0$ and $T_n \to \infty$ and f: $\mathbb{R}\times\Theta \to \mathbb{R}$. Moreover we suppose that $\exists c$: $|f(x,\theta)| \le c(1+|x|^c)$ and that $\beta \in (\frac{1}{4},\frac{1}{2})$.*
Then, $\forall \theta \in \Theta$,

1. $$\frac{1}{T_n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)\left(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i})\right)^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i}) 1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} f(x,\theta)a^2(x,\sigma_0)\pi(dx).$$

2. $$\frac{1}{n}\sum_{i=0}^{n-1} \frac{f(X_{t_i},\theta)}{\Delta_{n,i}}\left(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i})\right)^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i}) 1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} f(x,\theta)a^2(x,\sigma_0)\pi(dx).$$

The proof relies on the following lemma. In the sequel we will denote $\mathbb{E}_i[.]$ for $\mathbb{E}[.|\mathcal{F}_{t_i}]$, where $(\mathcal{F}_s)_s$ is the filtration defined in Lemma 1.

Lemma 3. *Suppose that Assumptions 1 to 4 hold. Moreover we suppose that $\beta \in (\frac{1}{4},\frac{1}{2})$. Then*

$$1.\mathbb{E}_i[(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2 \varphi^2_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] = \Delta_{n,i}a^2(X_{t_i},\sigma_0) + R(\theta,\Delta_{n,i}^{1+\beta},X_{t_i}), \tag{23}$$

$$2.\mathbb{E}_i[(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^4 \varphi^4_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] = 3\Delta_{n,i}^2 a^4(X_{t_i},\sigma_0) + R(\theta,\Delta_{n,i}^{\frac{7}{4}+\beta},X_{t_i}), \tag{24}$$

$$3.\textit{For } k \geq 1, \quad |\mathbb{E}_i[(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))\,\varphi^k_{\Delta^\beta_{n,i}}(X_{t_{i+1}} - X_{t_i})]| \leq R(\theta, \Delta_{n,i}, X_{t_i}), \tag{25}$$

$$4.\textit{For } k \geq 2,\ \forall k' > 0, \quad \mathbb{E}_i[|X_{t_{i+1}} - m(\mu,\sigma,X_{t_i})|^k\,|\varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}} - X_{t_i})|^{k'}] \leq R(\theta, \Delta_{n,i}^{\frac{k}{2}\wedge(1+\beta k)}, X_{t_i}). \tag{26}$$

$$5.\forall k' > 0, \quad \mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^3|\varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}} - X_{t_i})|^{k'}] = R(\theta_0, \Delta_{n,i}^{\frac{4}{3}+\beta}, X_{t_i}).$$

We observe that the first and the second points here above are particular cases of the the fourth one, in which we get some better estimation. In particular, we can identify in detail the main term.
Concerning the fifth point, instead, we remark that for $k = 3$ we don't have the main contribution of the Brownian part anymore, which gave us the rest function of size $\Delta_{n,i}^{\frac{k}{2}}$ in (26). In this case the main term of the development is given by the square of the Brownian integral times the jump part, which magnitude can be estimated by $\frac{4}{3} + \beta$.

In next lemma we consider the derivatives of φ, getting an improvement of the estimations here above. It relies on the fact that, from the definition we gave of such a function, we know its derivatives are different from zero only if the increments of our process are smaller than $2\Delta^\beta_{n,i}$ (as it was for φ) and bigger than $\Delta^\beta_{n,i}$ (extra bound that we did not get using φ). Having therefore no longer only an upper bound but also a lower bound for $X_{t_{i+1}} - X_{t_i}$, it is now possible to prove a better version of (26):

Lemma 4. *Suppose that Assumptions A1-A5 A7 and Ad hold. Then $\forall p \geq 1$, $\forall k \geq 1$ and $\forall r > 0$,*

$$\mathbb{E}_i[|X^\theta_{t_{i+1}} - m(\mu,\sigma,X_{t_i})|^p|\varphi^{(k)}_{\Delta^\beta_{n,i}}(X^\theta_{t_{i+1}} - X^\theta_{t_i})|^r] \leq R(\theta, h^{1+\beta p}, X_{t_i}).$$

Considering only the jump part, the following result holds:

Lemma 5. *Suppose that Assumptions A1-A4 holds. Then, $\forall q \geq 1$ we have*

$$\mathbb{E}_i[|\Delta X_i^J \varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|^q] = R(\theta_0, \Delta_{n,i}^{(1+\beta q)\wedge q}, X_{t_i}),$$

where $\Delta X_i^J := \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}} z\gamma(X_{s^-})\tilde{\mu}(ds,dz)$.

Other estimation about the expected value of the jump part in the presence of an indicator function which is 0 if the increments are bigger than $c\Delta^\beta_{n,i}$ are gathered in Lemma 4 of [3].

Using the lemmas stated here above, it is possible to prove the following proposition, that will be proved in Section A.1 and which is useful to show the tightness of the contrast function.

Proposition 7. *Suppose that Assumptions 1 to 4 and A_{Step} hold, $\Delta_n \to 0$ and $T_n \to \infty$ and $g_{i,n}$ is a differentiable function $\mathbb{R} \times \Theta \to \mathbb{R}$ such that $|g_{i,n}(x,\theta)| \leq c(1 + |x|^c)$ and, for $\vartheta = \mu$ and $\vartheta = \sigma$, $|\partial_\vartheta g_{i,n}(x,\theta)| \leq c(1 + |x|^c)$. We define*

$$S_n(\theta) := \frac{1}{T_n}\sum_{i=0}^{n-1}(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))\varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}} - X_{t_i})g_{i,n}(X_{t_i},\theta).$$

Then $S_n(\theta)$ is tight in $(C(\Theta), \|.\|_\infty)$.

7.2 Derivatives of m and m_2

We now state some propositions which concern the derivatives of m and m_2 that will be useful in the sequel. We introduce the following notation for the derivative operators $\partial_\vartheta := \frac{\partial}{\partial\vartheta}$, for $\vartheta = \mu$ and $\vartheta = \sigma$.

Proposition 8. *Suppose that Assumptions A1-A5 A7 and Ad hold. Then, for $|x| \leq \Delta_{n,i}^{-k_0}$ and $\forall \epsilon > 0$, we have*

$$1.\ \partial_\mu m(\mu,\sigma,X_{t_i}) = \Delta_{n,i}\partial_\mu b(X_{t_i},\mu) + R(\theta, \Delta_{n,i}^{\frac{5}{2}-\beta-\epsilon}, X_{t_i}),$$

$$2.\ |\partial_\sigma m(\mu,\sigma,X_{t_i})| \leq R(\theta, \Delta_{n,i}, X_{t_i}),$$

$$3.\ |\partial_\mu m_2(\mu,\sigma,X_{t_i})| \leq R(\theta, \Delta^2_{n,i}, X_{t_i}),$$

$$4.\ \partial_\sigma m_2(\mu,\sigma,X_{t_i}) = 2\Delta_{n,i}\partial_\sigma a(X_{t_i},\sigma)a(X_{t_i},\sigma) + R(\theta, \Delta_{n,i}^{1+\beta}, X_{t_i}).$$

We denote $\partial^2_\vartheta := \frac{\partial^2}{\partial\vartheta^2}$ and $\partial^2_{\mu\sigma} := \frac{\partial^2}{\partial\mu\partial\sigma}$. Estimation on the second derivatives are gathered in the following proposition:

Proposition 9. *Suppose that Assumptions A1 - A5, A7 and Ad hold. Then*

$$|\partial^2_{\mu\sigma} m(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}), \qquad |\partial^2_{\sigma} m(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i},X_{t_i}), \tag{27}$$

$$\partial^2_{\mu} m(\mu,\sigma,X_{t_i}) = \Delta_{n,i}\partial^2_{\mu} b(\mu,X_{t_i}) + R(\theta,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}), \tag{28}$$

$$|\partial^2_{\mu\sigma} m_2(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i}^{2},X_{t_i}), \qquad |\partial^2_{\mu} m_2(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i}^{2},X_{t_i}), \tag{29}$$

$$\partial^2_{\sigma} m_2(\mu,\sigma,X_{t_i}) = 2\Delta_{n,i}\partial_\sigma a(\sigma,X_{t_i})a(\sigma,X_{t_i}) + R(\theta,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}). \tag{30}$$

Deriving once again, the orders do not get worse. Indeed, the following estimations hold.

Proposition 10. *Suppose that Assumptions A1 - A5, A7 and Ad hold. Then*

1. $|\partial^3_{\mu} m_2(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i}^{2},X_{t_i})$;
2. $|\partial^3_{\sigma\mu\sigma} m_2(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i}^{2},X_{t_i})$,
3. $|\partial^3_{\mu\mu\sigma} m_2(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i}^{2},X_{t_i})$;
4. $|\partial^3_{\sigma} m_2(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i},X_{t_i})$,
5. $|\partial^3_{\mu} m(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i},X_{t_i})$;
6. $|\partial^3_{\sigma\mu\sigma} m(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i}^{\frac{3}{2}},X_{t_i})$,
7. $|\partial^3_{\mu\mu\sigma} m(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i}^{\frac{3}{2}},X_{t_i})$;
8. $|\partial^3_{\sigma} m(\mu,\sigma,X_{t_i})| \le R(\theta,\Delta_{n,i},X_{t_i})$.

The notation here above proposed for the third derivatives is the natural extension of the one introduced in Proposition 9 for the second derivatives.
Propositions 8, 9 and 10 will be proved in the appendix.

8 Proof of main results

We first of all study the asymptotic behaviour of the contrast, from which we find the consistency of our estimator.
We underline that, to get the consistency of the drift parameter, the normalization of the contrast function is different than the normalization we use to find the consistency of $\hat{\sigma}_n$. Even if it doesn't seem a natural choice, it works well on the basis of Proposition 6.

8.1 Contrast's convergence

To prove the contrast convergences, the development (10) of m_2 will be useful. We have shown in [2] (see (16) also) that under Assumptions $(A1)$-$(A4)$ the following development of $m(\mu,\sigma,x)$ holds :

$$m(\mu,\sigma,x) = x + \Delta_{n,i} b(x,\mu) + R^J(\Delta_{n,i},x) + r_1(\mu,\sigma,x), \tag{31}$$

where $r_1(\mu,\sigma,x)$ is a particular $R(\theta,\Delta_{n,i}^{1+\delta},X_{t_i})$ function (with $\delta>0$) and $R^J(\Delta_{n,i},x) = -\Delta_{n,i}\int_{\mathbb{R}} z\gamma(x)[1-\varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)]F(z)dz$; the J underlines that it turns out from a jump term. It has the same properties of the function R defined in Section 4.2 but it does not depend on θ.
Let us now prove the consistency of $\hat{\theta}_n$. The first step are the following lemmas:

Lemma 6. *Suppose that A1 - A5, A_{Step} and Ad hold. Moreover we suppose that $\beta \in (\frac{1}{4},\frac{1}{2})$. Then*

$$\frac{1}{n}U_n(\mu,\sigma) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} [\frac{c(x,\sigma_0)}{c(x,\sigma)} + \log(c(x,\sigma))]\pi(dx), \tag{32}$$

where $c(x,\sigma) = a^2(x,\sigma)$ and π is the invariant distribution defined in Lemma 2.

Lemma 6 is useful to prove the consistency of $\hat{\sigma}_n$, while we will use next lemma to show the consistency of $\hat{\mu}_n$

Lemma 7. *Suppose that A1 - A5, A_{Step} and Ad hold. Moreover we suppose that $\beta \in (\frac{1}{4},\frac{1}{2})$ and that $2\delta_1 > 1$. Then*

$$\frac{1}{T_n}(U_n(\mu,\sigma)-U_n(\mu_0,\sigma)) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} \frac{(b(x,\mu_0)-b(x,\mu))^2}{c(x,\sigma)}\pi(dx) + \int_{\mathbb{R}} [r(\mu,\sigma,x)-r(\mu_0,\sigma,x)](1-\frac{c(x,\sigma_0)}{c(x,\sigma)})\pi(dx), \tag{33}$$

where $r(\mu,\sigma,x)$ is the particular $R(\theta,1,x)$ function who turns out from the development (10) of m_2.

8.1.1 Proof of Lemma 6

Proof. We first of all observe that by the equation (10) we have, for $|X_{t_i}| \leq \Delta_{n,i}^{-k}$,

$$\frac{1}{m_2(\mu,\sigma,X_{t_i})} = \frac{1}{\Delta_{n,i}c(X_{t_i},\sigma)(1+\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}+\Delta_{n,i}r(\mu,\sigma,X_{t_i})+R(\theta,\Delta_{n,i}^{1+\delta_2},X_{t_i}))} =$$

$$= \frac{1}{\Delta_{n,i}c(X_{t_i},\sigma)}(1-\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}-\Delta_{n,i}r(\mu,\sigma,X_{t_i})+R(\theta,\Delta_{n,i}^{2\delta_1\wedge(1+\delta_2)\wedge 2},X_{t_i})). \tag{34}$$

In the sequel we will just write $\bar{r}$ for $2\delta_1\wedge(1+\delta_2)\wedge 2$. We observe that, as a consequence of the Assumption *Ad*, we have for both $\vartheta=\mu$ and $\vartheta=\sigma$, $|\partial_\vartheta R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i})| \leq R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i})$, where the two rest functions are not necessarily the same.
Similarly,

$$\log(\frac{m_2(\mu,\sigma,X_{t_i})}{\Delta_{n,i}}) = \log(c(X_{t_i},\sigma))+\log(1+\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}+\Delta_{n,i}r(\mu,\sigma,X_{t_i})+R(\theta,\Delta_{n,i}^{1+\delta_2},X_{t_i})) =$$

$$= \log(c(X_{t_i},\sigma))+\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}+\Delta_{n,i}r(\mu,\sigma,X_{t_i})+R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i}). \tag{35}$$

Using both (34) and (35) in the definition of $U_n(\mu,\sigma)$, we have to show that

$$\frac{1}{n}\sum_{i=0}^{n-1}\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}(1-\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}-\Delta_{n,i}r(\mu,\sigma,X_{t_i})+R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i}))+$$

$$+\frac{1}{n}\sum_{i=0}^{n-1}(\log(c(X_{t_i},\sigma))+\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}+\Delta_{n,i}r(\mu,\sigma,X_{t_i})+R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}} =: \sum_{j=1}^{8} I_j^n$$

converges to the right hand side of (32). We know that $I_1^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}}(\frac{c(x,\sigma_0)}{c(x,\sigma)})\pi(dx)$ because of Proposition 6. Using the third point of Proposition 5, $I_5^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}}\log(c(x,\sigma))\pi(dx)$. All the other terms converge to zero in norm 1 and so in probability. Indeed, passing through the conditional expectation and using the first point of Lemma 3 we have

$$\mathbb{E}[|I_2^n|] \leq \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|\frac{\Delta_{n,i}^{\delta_1}r(X_{t_i})}{\Delta_{n,i}c^2(X_{t_i},\sigma)}\mathbb{E}_i[(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})]1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}|] \leq$$

$$\leq \frac{\Delta_n^{\delta_1}}{n}\sum_{i=0}^{n-1}\mathbb{E}[|R(\theta,1,X_{t_i})|] \leq c\Delta_n^{\delta_1},$$

reminding that $r(X_{t_i})$ is a function $R(\theta,1,X_{t_i})$ by its definition and having used the property (9) on R, its polynomial growth and the third point of Lemma 2. In the same way we obtain

$$\mathbb{E}[|I_3^n|] \leq c\Delta_n \qquad \text{and} \qquad \mathbb{E}[|I_4^n|] \leq c\Delta_n^{\bar{r}},$$

that goes to zero since $\bar{r}=2\delta_1\wedge(1+\delta_2)\wedge 2$ is always positive.
Concerning I_6^n, as a consequence of the definition of $r(x)$ and the fact that $\Delta_{n,i}\leq\Delta_n$ we have again

$$\mathbb{E}[|I_6^n|] \leq \frac{\Delta_n^{\delta_1}}{n}\sum_{i=0}^{n-1}\mathbb{E}[|R(\theta,1,X_{t_i})|] \leq c\Delta_n^{\delta_1},$$

which converges to zero for $n\to\infty$. Again, acting in the same way we have

$$\mathbb{E}[|I_7^n|] \leq c\Delta_n \qquad \text{and} \qquad \mathbb{E}[|I_8^n|] \leq c\Delta_n^{\bar{r}}.$$

Convergence (32) follows. □

8.1.2 Proof of Lemma 7

Proof. Using again (34) and (35) we have that

$$\frac{1}{T_n}(U_n(\mu,\sigma)-U_n(\mu_0,\sigma)) = \frac{1}{T_n}\sum_{i=0}^{n-1}[\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}+$$

$$+\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}[\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}+$$

$$+\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}r(\mu,\sigma,X_{t_i})[1-\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}+$$

$$-\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}r(\mu_0,\sigma,X_{t_i})[1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}+$$

$$+\frac{1}{T_n}\sum_{i=0}^{n-1}[R((\mu,\sigma),\Delta_{n,i}^{\bar{r}},X_{t_i})+R((\mu,\sigma),\Delta_{n,i}^{\bar{r}},X_{t_i})\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}+$$

$$+\frac{1}{T_n}\sum_{i=0}^{n-1}[R((\mu_0,\sigma),\Delta_{n,i}^{\bar{r}},X_{t_i})+R((\mu_0,\sigma),\Delta_{n,i}^{\bar{r}},X_{t_i})\frac{(X_{t_{i+1}}-m(\mu_0,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}=:\sum_{j=1}^{6}I_j^n,$$

where we have introduced the notation $\Delta_iX := X_{t_{i+1}} - X_{t_i}$ and we recall that $\bar{r} = 2\delta_1 \wedge (1+\delta_2) \wedge 2$.
We have already proved in Lemma 4 of [2] that $I_1^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} \frac{(b(x,\mu_0)-b(x,\mu))^2}{c(x,\sigma)}\pi(dx)$. We observe that I_2^n differs from I_1^n only from the presence of $\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}$ and so, since δ_1 is positive, acting exactly as we did in order to prove the convergence of I_1^n it is possible to show that the added $\Delta_{n,i}^{\delta_1}$ make I_2^n converge to zero in probability.
Concerning I_3^n,

$$\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}r(\mu,\sigma,X_{t_i})\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}\xrightarrow{\mathbb{P}}\int_{\mathbb{R}}r(\mu,\sigma,x)\pi(dx) \tag{36}$$

as a consequence of the second point of Proposition 5. Moreover, using the third point of Proposition 6, we have that

$$\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}r(\mu,\sigma,X_{t_i})\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}\xrightarrow{\mathbb{P}}\int_{\mathbb{R}}r(\mu,\sigma,x)\frac{c(x,\sigma_0)}{c(x,\sigma)}\pi(dx). \tag{37}$$

From (36) and (37) it follows

$$I_3^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} r(\mu,\sigma,x)[1-\frac{c(x,\sigma_0)}{c(x,\sigma)}]\pi(dx).$$

Acting on I_4^n exactly like we did on I_3^n we get

$$I_4^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} r(\mu_0,\sigma,x)[1-\frac{c(x,\sigma_0)}{c(x,\sigma)}]\pi(dx).$$

Concerning I_5^n, it is

$$\frac{1}{T_n}\sum_{i=0}^{n-1}R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i})\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}\leq\Delta_n^{\bar{r}-1}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,1,X_{t_i})\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX),$$

which converges to zero in norm 1 and so in probability because of the boundedness of φ, the polynomial growth of R, the fact that $\frac{1}{T_n} = O(\frac{1}{n\Delta_n})$ and that $r-1$ is always positive since we have assumed $2\delta_1 > 1$. Moreover, passing through the conditional expectation and using the first point of Lemma 3 we have that

$$\frac{1}{T_n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i})\mathbb{E}_i[\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_iX)]1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}]\leq\Delta_n^{\bar{r}-1}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta,1,X_{t_i})]\leq c\Delta_n^{\bar{r}-1}.$$

We have therefore proved that the second part of I_5^n converges to 0 in norm 1 and therefore in probability. It follows $I_5^n \xrightarrow{\mathbb{P}} 0$ and, acting exactly in the same way, we have also $I_6^n \xrightarrow{\mathbb{P}} 0$. It yields (33). □

8.2 Consistency of the estimator.

In order to prove the consistency of $\hat{\theta}_n$, we need that the convergences (32) and (33) take place in probability uniformly in both the parameters, we want therefore to show the uniformity of the convergence in θ.
We regard $\frac{U_n(\mu,\sigma)}{n}$ and $S_n(\mu,\sigma) := \frac{1}{T_n}(U_n(\mu,\sigma) - U_n(\mu_0,\sigma))$ as random elements taking values in $(C(\Theta), \|.\|_\infty)$. It suffices to prove the tightness of these sequences; to do it we need some estimations for the derivatives of m and m_2 with respect to both the parameters, which are stated in Proposition 8, that will be proved in the Appendix. Such a proposition will be also useful to study the asymptotic behavior of the derivatives of the contrast function. We observe that, as a consequence of (31) and Proposition 8, for $\vartheta = \mu$ or $\vartheta = \sigma$ it is $|\partial_\vartheta r_1(\mu,\sigma,x)| \leq R(\theta, \Delta_{n,i}, X_{t_i})$.

Lemma 8. *Suppose that Assumption A1-A5, Ad, A_{Step} and A7 are satisfied. Then, the sequence $\frac{U_n(\mu,\sigma)}{n}$ is tight in $(C(\Theta), \|.\|_\infty)$.*

Proof. The tightness is implied by $\sup_n \frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma} |\partial_\vartheta U_n(\mu,\sigma)|] < \infty$ (see Corollary B.1 in [33]), for $\vartheta = \mu$ and $\vartheta = \sigma$. It is

$$\partial_\vartheta U_n(\mu,\sigma) = \sum_{i=0}^{n-1}[\frac{-2\partial_\vartheta m(\mu,\sigma,X_{t_i})(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))}{m_2(\mu,\sigma,X_{t_i})} - \frac{\partial_\vartheta m_2(\mu,\sigma,X_{t_i})(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2}{m_2^2(\mu,\sigma,X_{t_i})} + \\ + \frac{\partial_\vartheta m_2(\mu,\sigma,X_{t_i})}{m_2(\mu,\sigma,X_{t_i})}]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}}. \tag{38}$$

Using the first and the third point of Proposition 8 and the development (10) of m_2 it follows

$$\mathbb{E}[\sup_{\mu,\sigma} |\partial_\mu U_n(\mu,\sigma)|] \leq \sum_{i=0}^{n-1} \mathbb{E}[\sup_{\mu,\sigma} |R(\theta,1,X_{t_i})(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|1_{i,n}] + \\ + \sum_{i=0}^{n-1} \mathbb{E}[\sup_{\mu,\sigma} |R(\theta,1,X_{t_i})(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|1_{i,n}] + \sum_{i=0}^{n-1} \mathbb{E}[\sup_{\mu,\sigma} |R(\theta,\Delta_{n,i},X_{t_i})|1_{i,n}], \tag{39}$$

where we have used $1_{i,n}$ instead of $1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}}$ to shorten the notation.
We observe that

$$\mathbb{E}[\sup_{\mu,\sigma} |R(\theta,1,X_{t_i})(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|1_{i,n}] \leq \\ \leq \mathbb{E}[(\sup_{\mu,\sigma} |R(\theta,1,X_{t_i})|)(\sup_{\mu,\sigma} |(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|)1_{i,n}] \leq \\ \leq \mathbb{E}[(\sup_{\mu,\sigma} |R(\theta,1,X_{t_i})|1_{i,n})(|(X_{t_{i+1}} - m(\mu_0,\sigma,X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|)] + \\ + c\mathbb{E}[(\sup_{\mu,\sigma} |R(\theta,1,X_{t_i})|)(\sup_{\mu,\sigma} |m(\mu,\sigma,X_{t_i}) - m(\mu_0,\sigma,X_{t_i}))|)1_{i,n}]. \tag{40}$$

We can now use Cauchy-Schwartz inequality and (23) in Lemma 3 on the first, while on the second we use the development (31) of m getting that (40) is upper bounded by

$$c\mathbb{E}[R(\theta,\Delta_{n,i},X_{t_i})]^{\frac{1}{2}} + c\mathbb{E}[(\sup_{\mu,\sigma} |R(\theta,1,X_{t_i})|)(\sup_{\mu,\sigma} |\Delta_{n,i}(b(X_{t_i},\mu) - b(X_{t_i},\mu_0)) + r_1(\mu,\sigma,X_{t_i}) - r_1(\mu_0,\sigma,X_{t_i})|)] \leq \\ \leq c\Delta_n^{\frac{1}{2}} + c\mathbb{E}[\sup_{\mu,\sigma} |R(\theta,\Delta_{n,i},X_{t_i})|] \leq c\Delta_n^{\frac{1}{2}} + c\Delta_n \leq c\Delta_n^{\frac{1}{2}}, \tag{41}$$

where we have also used the boundedness of φ, the fact that R has polynomial growth uniformly in θ and the third point of Lemma 2 to say that our process has finite moments.
In the same way,

$$\mathbb{E}[\sup_{\mu,\sigma} |R(\theta,1,X_{t_i})(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|1_{i,n}] \leq \mathbb{E}[\sup_{\mu,\sigma} |R(\theta,\Delta_{n,i},X_{t_i})|] + \mathbb{E}[\sup_{\mu,\sigma} |R(\theta,\Delta_{n,i}^2,X_{t_i})|] \leq c\Delta_n. \tag{42}$$

Replacing (41) and (42) in (39) it follows

$$\sup_n \frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma} |\partial_\mu U_n(\mu,\sigma)|] \leq c\Delta_n^{\frac{1}{2}} \leq c < \infty.$$

We can act in the same way on $\partial_\sigma U_n(\mu,\sigma)$. Considering this time the second and the fourth point of Proposition 8 and still using the development (10) of m_2 and (23) in Lemma 3 it follows

$$\sup_n \frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\sigma U_n(\mu,\sigma)|] \leq \sup_n \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,\Delta_{n,i}^{\frac{1}{2}},X_{t_i})+R(\theta,1,X_{t_i})|] \leq c<\infty.$$

The tightness is therefore proved. □

Lemma 9. *Suppose that Assumption A1-A5, A7, A_{Step} and Ad are satisfied. We suppose moreover that δ_1 in the Assumption Ad is such that $2\delta_1>1$. Then, the sequence $S_n(\mu,\sigma)=\frac{1}{T_n}(U_n(\mu,\sigma)-U_n(\mu_0,\sigma))$ is tight in $(C(\Theta),\|.\|_\infty)$.*

Proof. We take again the notation used in the proof of Lemma 7, for which $S_n(\mu,\sigma)=\sum_{j=1}^6 I_j^n$. Since the sum of tight sequences is still tight, we will proceed showing that they are all tight. We start with I_3^n; acting as we did in Lemma 8, we prove that $\sup_n \mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta I_3^n|]<\infty$. We observe that, for $\vartheta=\mu$ and $\vartheta=\sigma$,

$$\partial_\vartheta I_3^n = \frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}[\partial_\vartheta r(\mu,\sigma,X_{t_i})(1-\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)})+$$

$$-r(\mu,\sigma,X_{t_i})\partial_\vartheta(\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)})]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}} =: I_{3,1}^n + I_{3,2}^n.$$

On $I_{3,1}^n$ we use the first point of Lemma 3 and that $|\partial_\vartheta r(\mu,\sigma,X_{t_i})|\leq R(\theta,1,X_{t_i})$ as stated in Assumption Ad to get

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma}|I_{3,1}^n|] \leq c+\sup_n\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[\mathbb{E}_i[\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|1_{i,n}]] \leq c \tag{43}$$

where we have used the polynomial growth of R, the third point of Lemma 2 and (42) and the notation $1_{i,n}$ instead of $1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}$.
Concerning $I_{3,2}^n$, the derivatives of $\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}$ with respect to μ and σ are different but in both cases they are upper bounded, using the first and the second point of Proposition 8, by $|R(\theta,1,X_{t_i})(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|$. We can therefore use (40) and (41), getting

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma}|I_{3,2}^n|] \leq c\sup_n \Delta_n^{\frac{1}{2}} \leq c<\infty. \tag{44}$$

From (43) and (44) it follows the tightness of I_3^n. Acting exactly in the same way on I_4^n it is clear it is tight too. Concerning I_5^n and I_6^n, recalling that the function $R(\theta,\Delta_{n,i}^{\bar r},X_{t_i})$ turns out from (34) and it is such that its derivatives with respect to both the parameters remains of the same order, we observe it is possible to act like we did on I_3^n getting

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma}|I_5^n|] \leq c\Delta_n^{\bar r-1}+c\Delta_n^{\bar r-\frac{1}{2}}<\infty,$$

since we have chosen $2\delta_1>1$ and so $\bar r-1$ is positive. Clearly the same estimation hold for I_6^n.
We now prove that I_1^n is tight. To do it we observe that, using the development (31) and the dynamic (3) of the process X we have

$$X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}) = \int_{t_i}^{t_{i+1}} b(X_s,\mu_0)ds+\int_{t_i}^{t_{i+1}} a(\sigma_0,X_s)dW_s+\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}\setminus\{0\}}\gamma(X_{s^-})z\tilde\mu(ds,dz)+ \tag{45}$$

$$-R^J(\Delta_{n,i},X_{t_i})-\Delta_{n,i}b(X_{t_i},\mu)-r_1(\mu,\sigma,X_{t_i}).$$

It is worth noting that only the last two terms here above depend on μ and so replacing (45) in I_1^n some terms are deleted by compensation. Therefore we can define

$$I_{1,1}^n := \frac{1}{T_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}}{\Delta_{n,i}c(\sigma,X_{t_i})}[\Delta_{n,i}^2(b^2(X_{t_i},\mu)-b^2(X_{t_i},\mu_0))+r_1^2(\mu,\sigma,X_{t_i})-r_1^2(\mu_0,\sigma,X_{t_i})+$$

$$+2\Delta_{n,i}b(X_{t_i},\mu)r_1(\mu,\sigma,X_{t_i})-2\Delta_{n,i}b(X_{t_i},\mu_0)r_1(\mu_0,\sigma,X_{t_i})+2[\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds+\Delta X_i^J+$$

$$-R^J(\Delta_{n,i},X_{t_i})][\Delta_{n,i}(b(X_{t_i},\mu)-b(X_{t_i},\mu_0))+r_1(\mu,\sigma,X_{t_i})-r_1(\mu_0,\sigma,X_{t_i})],$$

where we have denoted by ΔX_i^J the jump part in ΔX_i, that is $\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}\setminus\{0\}}\gamma(X_{s^-})z\tilde{\mu}(ds,dz)$. Moreover we define

$$I^n_{1,2}:=\frac{1}{T_n}\sum_{i=0}^{n-1}\frac{2(b(X_{t_i},\mu)-b(X_{t_i},\mu_0))\int_{t_i}^{t_{i+1}}a(\sigma_0,X_s)dW_s}{c(\sigma,X_{t_i})}\varphi_{\Delta^\beta_{n,i}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}},$$

$$I^n_{1,3}:=\frac{1}{T_n}\sum_{i=0}^{n-1}\frac{2(r_1(\mu,\sigma,X_{t_i})-r_1(\mu_0,\sigma,X_{t_i}))\int_{t_i}^{t_{i+1}}a(\sigma_0,X_s)dW_s}{\Delta_{n,i}c(\sigma,X_{t_i})}\varphi_{\Delta^\beta_{n,i}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}.$$

It is $I^n_1=I^n_{1,1}+I^n_{1,2}+I^n_{1,3}$. We are going to prove that $I^n_{1,1}$ is tight showing that the expected value of the derivatives is bounded, like we have already done. On $I^n_{1,2}$ and $I^n_{1,3}$ we will use instead the Kolmogorov criterion for which, if for some positive constant H independent of n and for $m\geq r>2$, S_n is a sequence such that

$$\mathbb{E}[(S_n(\theta))^m]\leq H\qquad\forall\theta\in\Theta, \tag{46}$$

$$\mathbb{E}[(S_n(\theta_1)-S_n(\theta_2))^m]\leq H|\mu_1-\mu_2|^r+H|\sigma_1-\sigma_2|^r\qquad\forall\theta_1,\theta_2\in\Theta, \tag{47}$$

then S_n is tight.
Let us start considering $I^n_{1,1}$: we want to show that $\sup_n\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta I^n_{1,1}|]<\infty$. We observe it is

$$\partial_\mu I^n_{1,1}=\frac{1}{T_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta^\beta_{n,i}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}}{\Delta_{n,i}c(\sigma,X_{t_i})}[\Delta^2_{n,i}(2b\,\partial_\mu b)(X_{t_i},\mu)+(2r_1\partial_\mu r_1)(\mu,\sigma,X_{t_i})+2\Delta_{n,i}((\partial_\mu b)(X_{t_i},\mu)r_1(\mu,\sigma,X_{t_i})+$$

$$+b(X_{t_i},\mu)(\partial_\mu r_1)(\mu,\sigma,X_{t_i}))+2(\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds+\Delta X_i^J-R^J(\Delta_{n,i},X_{t_i}))(\Delta_{n,i}\partial_\mu b(X_{t_i},\mu)+\partial_\mu r_1(\mu,\sigma,X_{t_i}))];$$

$$\partial_\sigma I^n_{1,1}=\frac{1}{T_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta^\beta_{n,i}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}}{\Delta_{n,i}c(\sigma,X_{t_i})}[2r_1\partial_\sigma r_1(\mu,\sigma,X_{t_i})-2r_1\partial_\sigma r_1(\mu_0,\sigma,X_{t_i})+2\Delta_{n,i}(b(X_{t_i},\mu)\partial_\sigma r_1(\mu,\sigma,X_{t_i})+$$

$$-b(X_{t_i},\mu_0)\partial_\sigma r_1(\mu_0,\sigma,X_{t_i}))+2(\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds+\Delta X_i^J-R^J(\Delta_{n,i},X_{t_i}))(\partial_\sigma r_1(\mu_0,\sigma,X_{t_i})-\partial_\sigma r_1(\mu,\sigma,X_{t_i}))]+$$

$$-\frac{\partial_\sigma c(X_{t_i},\sigma)\varphi_{\Delta^\beta_{n,i}}(\Delta_iX)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}}{\Delta_{n,i}c^2(\sigma,X_{t_i})}[\Delta^2_{n,i}(b^2(X_{t_i},\mu)-b^2(X_{t_i},\mu_0))+r_1^2(\mu,\sigma,X_{t_i})-r_1^2(\mu_0,\sigma,X_{t_i})+$$

$$+2\Delta_{n,i}b(X_{t_i},\mu)r_1(\mu,\sigma,X_{t_i})-2\Delta_{n,i}b(X_{t_i},\mu_0)r_1(\mu_0,\sigma,X_{t_i})+2(\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds+\Delta X_i^J+$$

$$-R^J(\Delta_{n,i},X_{t_i}))(\Delta_{n,i}(b(X_{t_i},\mu)-b(X_{t_i},\mu_0))+r_1(\mu,\sigma,X_{t_i})-r_1(\mu_0,\sigma,X_{t_i}))].$$

Using the polynomial growth of b and recalling that r_1 is the particular $R(\theta,\Delta_{n,i}^{1+\delta},X_{t_i})$ function that turns out from the development (31) of m and it is such that $|\partial_\vartheta r_1(\mu,\sigma,X_{t_i})|\leq R(\theta,\Delta_{n,i},X_{t_i})$ as a consequence of the first two points of Proposition 8, we get

$$|\partial_\vartheta I^n_{1,1}|\leq\frac{c}{n\Delta_n}\sum_{i=0}^{n-1}|\varphi_{\Delta^\beta_{n,i}}(\Delta_iX)|[R(\theta,\Delta_{n,i},X_{t_i})+R(\theta,\Delta_{n,i}^{1+\delta},X_{t_i})+$$

$$+2(|\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds|+|\Delta X_i^J|+R^J(\Delta_{n,i},X_{t_i}))(R(\theta,1,X_{t_i})+R(\theta,\Delta_{n,i}^{\delta},X_{t_i}))].$$

Using Lemma 5, the boundedness of φ, the fact that $\frac{1}{T_n}=O(\frac{1}{n\Delta_n})$ and that $2\mathbb{E}_i[|\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds|]$ is a $R(\theta_0,\Delta_{n,i},X_{t_i})$, it follows

$$\sup_n\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta I^n_{1,1}|]\leq\sup_n(\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,\Delta_{n,i},X_{t_i})+(R(\theta_0,\Delta_{n,i},X_{t_i})+R^J(\Delta_{n,i},X_{t_i}))R(\theta,1,X_{t_i})|]+$$

$$+\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[(\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})|)\mathbb{E}_i[|\Delta X_i^J\varphi_{\Delta^\beta_{n,i}}(\Delta_iX)|]])\leq c,$$

where in the last inequality we have used Lemma 5 here above, the polynomial growth of R uniform in θ and the third point of Lemma 2. $I^n_{1,1}$ is therefore tight.

We now show that (46) and (47) hold on $I^n_{1,2}$. Indeed, using Burkholder and Jensen inequalities, we get $\mathbb{E}[|I^n_{1,2}(\theta_1) - I^n_{1,2}(\theta_2)|^m] \leq$

$$\leq \frac{c}{n^m\Delta_n^m} n^{\frac{m}{2}-1} \sum_{i=0}^{n-1} \mathbb{E}[|\frac{b(X_{t_i},\mu_1)-b(X_{t_i},\mu_0)}{c(\sigma_1,X_{t_i})} - \frac{b(X_{t_i},\mu_2)-b(X_{t_i},\mu_0)}{c(\sigma_2,X_{t_i})}|^m | \int_{t_i}^{t_{i+1}} a(\sigma_0,X_s)dW_s|^m |\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|^m]. \tag{48}$$

We observe that, as a consequence of the finite-increments theorem, we have

$$|\frac{b(X_{t_i},\mu_1)-b(X_{t_i},\mu_0)}{c(\sigma_1,X_{t_i})} - \frac{b(X_{t_i},\mu_2)-b(X_{t_i},\mu_0)}{c(\sigma_2,X_{t_i})}|^m \leq |\frac{\partial_\mu b(X_{t_i},\tilde{\mu})}{c(X_{t_i},\tilde{\sigma})}(\mu_1-\mu_2) + \tag{49}$$

$$-\frac{(b(X_{t_i},\tilde{\mu})-b(X_{t_i},\mu_0))\partial_\sigma c(X_{t_i},\tilde{\sigma})}{c(X_{t_i},\tilde{\sigma})}(\sigma_1-\sigma_2)|^m \leq R(\theta,1,X_{t_i})|\mu_1-\mu_2|^m + R(\theta,1,X_{t_i})|\sigma_1-\sigma_2|^m,$$

where actually the functions R are calculated in a point $\tilde{\theta} := (\tilde{\mu},\tilde{\sigma})$, with $\tilde{\mu} \in (\mu_1,\mu_2)$ and $\tilde{\sigma} \in (\sigma_1,\sigma_2)$ but, since the property (8) of R is uniform in θ, we have chosen to write it simply as $R(\theta,1,X_{t_i})$. Using also the boundedness of φ, we get that (48) is upper bounded by

$$\frac{cn^{\frac{m}{2}-1}}{n^m\Delta_n^m}\sum_{i=0}^{n-1}\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(\sigma_0,X_s)dW_s|^m R(\theta,1,X_{t_i})]|\mu_1-\mu_2)|^m + \mathbb{E}[|\int_{t_i}^{t_{i+1}} a(\sigma_0,X_s)dW_s|^m R(\theta,1,X_{t_i})]|\sigma_1-\sigma_2|^m.$$

Using Burkholder-Davis-Gundy inequality we have, $\forall p \geq 2$,

$$\mathbb{E}[(\int_{t_i}^{t_{i+1}} a(\sigma,X_s)dW_s)^p] \leq \mathbb{E}[(\int_{t_i}^{t_{i+1}} a^2(\sigma,X_s)ds)^{\frac{p}{2}}] \leq \mathbb{E}[R(\theta,\Delta_{n,i},X_{t_i})^{\frac{p}{2}}] = c\Delta_{n,i}^{\frac{p}{2}}, \tag{50}$$

where in the last inequality we have used the polynomial growth of a and the third point of Lemma 2. From Holder inequality and (50) it therefore follows

$$\mathbb{E}[|I^n_{1,2}(\theta_1) - I^n_{1,2}(\theta_2)|^m] \leq \frac{c}{(n\Delta_n)^{\frac{m}{2}}}|\mu_1-\mu_2|^m + \frac{c}{(n\Delta_n)^{\frac{m}{2}}}|\sigma_1-\sigma_2|^m \leq c|\mu_1-\mu_2|^m + c|\sigma_1-\sigma_2|^m,$$

where we have also used that $n\Delta_n \to \infty$ for $n \to \infty$. For $r := m$ (47) is proved.
Concerning (46), acting in the same way we get

$$\mathbb{E}[|I^n_{1,2}(\theta)|^m] \leq \frac{cn^{\frac{m}{2}-1}}{n^m\Delta_n^m}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta_1,1,X_{t_i})^m|\int_{t_i}^{t_{i+1}} a(\sigma_0,X_s)dW_s|^m|\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|^m] \leq \frac{c}{(n\Delta_n)^{\frac{m}{2}}} \leq c.$$

$I^n_{1,2}$ is hence tight. The tightness of $I^n_{1,3}$ is obtained acting exactly in the same way, remarking that

$$|\frac{r_1(\mu_0,\sigma_1,X_{t_i})-r_1(\mu_1,\sigma_1,X_{t_i})}{\Delta_{n,i}c(\sigma_1,X_{t_i})} - \frac{r_1(\mu_0,\sigma_2,X_{t_i})-r_1(\mu_2,\sigma_2,X_{t_i})}{\Delta_{n,i}c(\sigma_2,X_{t_i})}|^m \leq |\frac{\partial_\mu r_1(\tilde{\mu},\tilde{\sigma},X_{t_i})}{\Delta_{n,i}c(X_{t_i},\tilde{\sigma})}(\mu_1-\mu_2) +$$

$$+[\frac{\partial_\sigma r_1(\mu_0,\tilde{\sigma},X_{t_i})-\partial_\sigma r_1(\tilde{\mu},\tilde{\sigma},X_{t_i})}{\Delta_{n,i}c(\tilde{\sigma},X_{t_i})} - \frac{\partial_\sigma c(\tilde{\sigma},X_{t_i})(r_1(\mu_0,\tilde{\sigma},X_{t_i})-r_1(\tilde{\mu},\tilde{\sigma},X_{t_i}))}{\Delta_{n,i}c(\tilde{\sigma},X_{t_i})}](\sigma_1-\sigma_2)|^m \leq \tag{51}$$

$$\leq R(\theta,1,X_{t_i})|\mu_1-\mu_2|^m + R(\theta,1,X_{t_i})|\sigma_1-\sigma_2|^m,$$

as a consequence of the fact that $(r_1(\mu,\sigma,X_{t_i}))^m$ and $(\partial_\vartheta r_1(\mu,\sigma,X_{t_i}))^m$ are respectively upper bounded by $R(\theta,\Delta_{n,i}^{m(1+\delta)},X_{t_i})$ and $R(\theta,\Delta^m_{n,i},X_{t_i})$.
Concerning I^n_2, we act like we did on I^n_1. We still use (45) getting $I^n_{2,1}$, $I^n_{2,2}$ and $I^n_{2,3}$. We observe that, if we define s^n_j as $I^n_{1,j} =: \sum_{i=0}^{n-1} s^n_j$, then $I^n_{2,j} = \sum_{i=0}^{n-1}\Delta^\delta_{n,i}\frac{r(X_{t_i})}{c(\sigma,X_{t_i})}s^n_j$.
By the computation of $\partial_\mu I^n_{2,1}$ and $\partial_\sigma I^n_{2,1}$ it follows that

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta I^n_{2,1}|] \leq \sup_n(c\Delta_n^\delta + c\Delta_n^{\delta+\beta}) \leq c.$$

In order to prove that also $I^n_{2,2}$ and $I^n_{2,3}$ are tight we still use Kolmogorov criterion. From (49) and (50) it follows

$$\mathbb{E}[|I^n_{2,2}(\theta_1) - I^n_{2,2}(\theta_2)|^m] \leq c\frac{\Delta_n^{\delta m}}{(n\Delta_n)^{\frac{m}{2}}}|\mu_1-\mu_2|^m + c\frac{\Delta_n^{\delta m}}{(n\Delta_n)^{\frac{m}{2}}}|\sigma_1-\sigma_2|^m \leq c|\mu_1-\mu_2|^m + c|\sigma_1-\sigma_2|^m$$

and $\mathbb{E}[(I^n_{2,2}(\theta))^m] \leq c$.
The tightness of $I^n_{2,3}$ is obtained in the same way, through Kolmogorov criterion and (51).
The sequence S_n is therefore tight in $(C(\Theta),\|.\|_\infty)$, as we wanted. □

8.2.1 Proof of Theorem 1.

Proof. Let us begin with the consistency of $\hat{\sigma_n}$. An application of lemmas 6 and 8 yields

$$\frac{1}{n}U_n(\mu,\sigma) \xrightarrow{\mathbb{P}} U(\sigma,\sigma_0) := \int_{\mathbb{R}} [\frac{c(x,\sigma_0)}{c(x,\sigma)} + \log(c(x,\sigma))]\pi(dx) \tag{52}$$

uniformly in θ.
In order to prove that the uniform convergence here above implies the consistency of $\hat{\sigma_n}$, since the convergence in probability is equivalent to the existence, for any subsequence, of a subsequence converging almost surely, we will consider that the convergence in (52) is almost sure and prove that it implies that $\hat{\sigma}_n \to \sigma_0$ almost surely. For a fixed ω, thanks to the compactness of Θ, there exists a subsequence n_k such that $(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k})$ tends to a limit $\theta_\infty := (\mu_\infty, \sigma_\infty)$. Hence, (52) together with the continuity of $\sigma \mapsto U(\sigma, \sigma_0)$, implies

$$\frac{1}{n_k}U_{n_k}(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k})(\omega) \to U(\sigma_\infty, \sigma_0).$$

But, by the definition of our estimator $\hat{\theta}_n$,

$$\frac{1}{n_k}U_{n_k}(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k}) \leq \frac{1}{n_k}U_{n_k}(\hat{\mu}_{n_k}, \sigma_0).$$

So, using again the convergence (52), we get $U(\sigma_\infty, \sigma_0) \leq U(\sigma_0, \sigma_0)$. On the other hand, since for all $y > 0$, $y_0 > 0$ it is $\frac{y_0}{y} + \log(y) \geq 1 + \log(y_0)$ we deduce, using also the identifiability stated in Assumption A6 and Proposition 8.1 in Supplemental materials of [15], that $\sigma_\infty = \sigma_0$. We have proved that any convergent subsequence of $\hat{\sigma}_n$ tends to σ_0, hence $\hat{\sigma}_n \xrightarrow{\mathbb{P}} \sigma_0$ and we are done.
Concerning the consistency of $\hat{\mu}_n$, we have from Lemmas 7 and 9 that the convergence (33) holds uniformly in θ. In order to deduce the consistency of $\hat{\mu}_n$ the method is similar to the previous one. We know now that $(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k})$ tends to (μ_∞, σ_0), hence

$$\frac{1}{T_{n_k}}(U_{n_k}(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k}) - U_{n_k}(\mu_0, \hat{\sigma}_{n_k})) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} \frac{(b(x,\mu_0) - b(x,\mu_\infty))^2}{c(x,\sigma_0)}\pi(dx) \geq 0.$$

But $U_{n_k}(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k}) - U_{n_k}(\mu_0, \hat{\sigma}_{n_k}) \leq 0$ and so we conclude by A6, getting $\mu_\infty = \mu_0$ and therefore the consistency of $\hat{\mu}_n$. □

8.3 Asymptotic normality of the estimator.

The proof of the asymptotic normality goes along a classical route (see for instance Section 5a of [14]). We define the following notations:

$$M_n := \begin{pmatrix} \frac{1}{\sqrt{T_n}} & 0 \\ 0 & \frac{1}{\sqrt{n}} \end{pmatrix}.$$

Let

$$S_n := \begin{pmatrix} \sqrt{T_n}(\hat{\mu}_n - \mu_0) \\ \sqrt{n}(\hat{\sigma}_n - \sigma_0) \end{pmatrix}, \qquad L_n(\theta_0) := \begin{pmatrix} -\frac{1}{\sqrt{T_n}}\partial_\mu U_n(\mu_0, \sigma_0) \\ -\frac{1}{\sqrt{n}}\partial_\sigma U_n(\mu_0, \sigma_0) \end{pmatrix}$$

and

$$C_n(\theta) = \begin{pmatrix} \frac{1}{T_n}\frac{\partial^2}{\partial\mu^2}U_n(\mu,\sigma) & \frac{1}{\sqrt{nT_n}}\frac{\partial^2}{\partial\mu\sigma}U_n(\mu,\sigma) \\ \frac{1}{\sqrt{nT_n}}\frac{\partial^2}{\partial\mu\sigma}U_n(\mu,\sigma) & \frac{1}{n}\frac{\partial^2}{\partial\sigma^2}U_n(\mu,\sigma) \end{pmatrix}.$$

Then

$$M_n\nabla^2_\theta U_n(\mu,\sigma)M_n = C_n(\theta). \tag{53}$$

Now, by Taylor's formula,

$$\int_0^1 \nabla^2_\theta U_n(\theta_0 + u(\hat{\theta}_n - \theta_0))du \begin{pmatrix} \hat{\mu}_n - \mu_0 \\ \hat{\sigma}_n - \sigma_0 \end{pmatrix} = -\nabla_\theta U_n(\theta_0),$$

since $\nabla_\theta U_n(\hat{\theta}_n) = 0$. Then, using (53), we have

$$\int_0^1 C_n(\theta_0 + u(\hat{\theta}_n - \theta_0))du S_n = L_n(\theta_0). \tag{54}$$

We deduce from this equality that, in order to prove the asymptotic normality of $\hat{\theta}_n$ and hence to end the proof of Theorem 2, it is enough to prove the following lemmas:

Lemma 10. *Suppose that Assumptions A1-A8 and Ad hold. Then, as $n \to \infty$,*

$$L_n(\theta_0) \xrightarrow{d} L \sim N(0, K'),$$

where $K' = \begin{pmatrix} 4\int_{\mathbb{R}} (\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2 \pi(dx) & 0 \\ 0 & 8\int_{\mathbb{R}} (\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2 \pi(dx) \end{pmatrix}$.

Lemma 11. *Suppose that Assumptions A1-A8 and Ad hold. Then the following statements hold:*

$$1.\, C_n(\theta_0) \xrightarrow{\mathbb{P}} B = \begin{pmatrix} -2\int_{\mathbb{R}} (\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2 \pi(dx) & 0 \\ 0 & 4\int_{\mathbb{R}} (\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2 \pi(dx) \end{pmatrix},$$

$$2. \sup_{\{|\tilde{\theta}| \le \epsilon_n\}} |C_n(\theta_0 + \tilde{\theta}) - C_n(\theta_0)| \xrightarrow{\mathbb{P}} 0, \quad \textit{where } \epsilon_n \to 0.$$

8.3.1 Proof of Lemma 10.

Proof. As a consequence of a combination of Theorem 3.2 and Theorem 3.4 in [18] (c.f. also Section A.2 in the Appendix of [33]) we get the result if we prove that $L_n(\theta_0)$ is a triangular array of martingale increments such that, for a constant $r > 0$, the following convergences hold.
We define ζ_i and $\tilde{\zeta}_i$ such that $\partial_\mu U_n(\mu_0, \sigma_0) =: \sum_{i=0}^{n-1} \zeta_i(\theta_0)$ and $\partial_\sigma U_n(\mu_0, \sigma_0) =: \sum_{i=0}^{n-1} \tilde{\zeta}_i(\theta_0)$. Then it must be

$$\frac{1}{T_n}\sum_{i=0}^{n-1} \mathbb{E}_i[\zeta_i^2(\theta_0)] \xrightarrow{\mathbb{P}} 4\int_{\mathbb{R}} (\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2 \pi(dx) \qquad \frac{1}{(\sqrt{T_n})^{2+r}}\sum_{i=0}^{n-1} \mathbb{E}_i[\zeta_i^{2+r}(\theta_0)] \xrightarrow{\mathbb{P}} 0, \tag{55}$$

$$\frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}_i[\tilde{\zeta}_i^2(\theta_0)] \xrightarrow{\mathbb{P}} 8\int_{\mathbb{R}} (\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2 \pi(dx) \qquad \frac{1}{(\sqrt{n})^{2+r}}\sum_{i=0}^{n-1} \mathbb{E}_i[\tilde{\zeta}_i^{2+r}(\theta_0)] \xrightarrow{\mathbb{P}} 0, \tag{56}$$

$$\frac{1}{\sqrt{nT_n}}\sum_{i=0}^{n-1} |\mathbb{E}_i[\zeta_i(\theta_0)\tilde{\zeta}_i(\theta_0)]| \xrightarrow{\mathbb{P}} 0. \tag{57}$$

First of all we observe that $L_n(\theta_0)$ is a triangular array of martingale increments as a consequence of the definitions of m and m_2. Indeed, using (38), we clearly have

$$\mathbb{E}_i[\zeta_i(\theta_0)] = \frac{-2\partial_\mu m(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})}\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} +$$

$$+\frac{\partial_\mu m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})}\mathbb{E}_i[(1 - \frac{(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} = 0.$$

In the same way, computing the derivative with respect to σ we clearly have $\mathbb{E}_i[\tilde{\zeta}_i(\theta_0)] = 0$.
Concerning $\partial_\mu U_n$, using (34) we can see $\zeta_i(\theta_0)$ as

$$\frac{-2\partial_\mu m(\mu_0,\sigma_0,X_{t_i})}{\Delta_{n,i} c(X_{t_i},\sigma_0)}(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} + R_{i,n}(\theta_0) =: \hat{\zeta}_i(\theta_0) + R_{i,n}(\theta_0),$$

we have already proved in Lemma 6 of [2] the asymptotic normality of $\frac{1}{\sqrt{T_n}}\sum_{i=0}^{n-1} \hat{\zeta}_i(\theta_0)$ and, in particular, that convergences (55) hold with $\hat{\zeta}_i(\theta_0)$ instead of $\zeta_i(\theta_0)$.
In order to conclude the proof of (55), it is enough to have $\frac{1}{T_n}\sum_{i=0}^{n-1} \mathbb{E}_i[R_{i,n}^2(\theta_0)] \xrightarrow{\mathbb{P}} 0$ and $(\frac{1}{\sqrt{T_n}})^{2+r}\sum_{i=0}^{n-1} \mathbb{E}_i[R_{i,n}^{2+r}(\theta_0)] \xrightarrow{\mathbb{P}} 0$. It is

$$\frac{1}{T_n}\sum_{i=0}^{n-1} \mathbb{E}_i[R_{i,n}^2(\theta_0)] \le \frac{c}{n\Delta_n}\sum_{i=0}^{n-1} (\frac{\partial_\mu m(\mu_0,\sigma_0,X_{t_i})R(\theta,\Delta_{n,i}^{\delta_1\wedge 1}X_{t_i})}{\Delta_{n,i} c(X_{t_i},\sigma_0)})^2 \mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^2 \varphi_{\Delta_{n,i}^\beta}^2(\Delta_i X)] +$$

$$+\frac{c}{n\Delta_n}\sum_{i=0}^{n-1} (\frac{\partial_\mu m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^2 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} +$$

$$+\frac{c}{n\Delta_n}\sum_{i=0}^{n-1} (\frac{\partial_\mu m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^2 \frac{\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^4 \varphi_{\Delta_{n,i}^\beta}^2(\Delta_i X)]1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}}{m_2^2(\mu_0,\sigma_0,X_{t_i})}$$

As a consequence of the first and the third point of Proposition 8 and using first and second point of Lemma 3 it is upper bounded by

$$\frac{c}{n}\sum_{i=0}^{n-1} R(\theta_0, \Delta_{n,i}^{2\delta_1\wedge 2}, X_{t_i}) + \frac{c}{n\Delta_n}\sum_{i=0}^{n-1} R(\theta_0, \Delta_{n,i}^2, X_{t_i}),$$

which converges to zero in norm 1 and so in probability.
Acting in the same way, using this time the fourth point of Lemma 3 twice, for $k = (2+r)$ and $k = 2(2+r)$, it follows that also $(\frac{1}{\sqrt{T_n}})^{2+r}\sum_{i=0}^{n-1}\mathbb{E}_i[R_{i,n}^{2+r}(\theta_0)]$ goes to zero in probability.
Concerning the derivative of the contrast with respect to σ, it is

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[\tilde{\zeta}_i^2(\theta_0)] = \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[(A_i^2 + 2A_iB_i + B_i^2)\varphi^2_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}},$$

where we have defined

$$A_i := \frac{-2\partial_\sigma m(\mu_0,\sigma_0,X_{t_i})(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))}{m_2(\mu_0,\sigma_0,X_{t_i})}$$

and

$$B_i := (\frac{\partial_\sigma m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})(1 - \frac{(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})}).$$

By the development (10) of m_2, the second point of Proposition 8 and equation (23) in Lemma 3 we have

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[A_i^2\varphi^2_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}} \le \frac{1}{n}\sum_{i=0}^{n-1} R(\theta_0,\Delta_{n,i},X_{t_i}),$$

that goes to zero in norm 1 because of the property (9) of R, its polynomial growth and the third point of Lemma 2. The convergence to zero in probability follows.
On the mixed term we use the development (10) of m_2, the second and the fourth point of Proposition 8 to get respectively an upper bound on the derivatives with respect to σ of m and m_2 and the first and the fifth point of Lemma 3 to obtain the following:

$$\frac{1}{n}\sum_{i=0}^{n-1}|\mathbb{E}_i[2A_iB_i\varphi^2_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}| \le \frac{1}{n}\sum_{i=0}^{n-1}\big[\,|\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))\varphi^2_{\Delta_{n,i}^\beta}(\Delta_i X)]| +$$

$$+R(\theta_0,\Delta_{n,i}^{-1},X_{t_i})|\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^3\varphi^2_{\Delta_{n,i}^\beta}(\Delta_i X)]|\,\big] \le \frac{1}{n}\sum_{i=0}^{n-1}[R(\theta_0,\Delta_{n,i},X_{t_i}) + R(\theta_0,\Delta_{n,i}^{\frac{1}{3}+\beta},X_{t_i})].$$

We obtain that the mixed term converges to zero in probability for the same argument we gave for the convergence of A_i^2: because of the property (9) of R, its polynomial growth and the third point of Lemma 2 we get the convergence in norm 1 which implies the convergence in probability.
We now study the convergence of the term

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[B_i^2\varphi^2_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}.$$

Using the fourth point of Proposition 8, the development (10) of m_2 and (34), it is

$$\frac{1}{n}\sum_{i=0}^{n-1}[(\frac{2\partial_\sigma a(X_{t_i},\sigma_0)a(X_{t_i},\sigma_0)}{c(X_{t_i},\sigma_0)})^2 + R(\theta_0,\Delta_{n,i}^{\beta\wedge\delta_1},X_{t_i})]\mathbb{E}_i[(1 - \frac{(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})^2\varphi^2_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}. \tag{58}$$

We now need the following lemma:

Lemma 12. *Suppose that Assumptions A1-A4 hold. Then, $\forall q \ge 1$,*

$$\mathbb{E}_i[|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|^q] = 1 + R(\theta,\Delta_{n,i},X_{t_i}).$$

Proof. Lemma 12.
We can see $\mathbb{E}_i[|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|^q]$ as $1 + \mathbb{E}_i[|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|^q - 1]$.
Because of the definition of φ,the expected value here above is different from zero only if $|\Delta_i X| \ge \Delta_{n,i}^\beta$. Hence, using (71) it is

$$\mathbb{E}_i[|\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|^q - 1] \le c\mathbb{E}_i[1_{\{|\Delta_i X|\ge\Delta_{n,i}^\beta\}}] \le R(\theta,\Delta_{n,i},X_{t_i}).$$

□

Using the lemma here above, still the development (10) of m_2 and (23) and (24) in Lemma 3 we have

$$\mathbb{E}_i[(1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})^2\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)] = 1+R(\theta_0,\Delta_{n,i}^{1\wedge\delta_1},X_{t_i})+3\frac{a^4(X_{t_i},\sigma_0)}{c^2(X_{t_i},\sigma_0)}-2\frac{a^2(X_{t_i},\sigma_0)}{c(X_{t_i},\sigma_0)}+$$

$$+R(\theta_0,\Delta_{n,i}^{\delta_1\wedge(\beta-\frac{1}{4})},X_{t_i}) = 2+R(\theta_0,\Delta_{n,i}^{\delta_1\wedge(\beta-\frac{1}{4})},X_{t_i}),$$

we remind that $c(x,\sigma)=a^2(x,\sigma)$. Replacing the last equation in (58) we get

$$\frac{1}{n}\sum_{i=0}^{n-1}2\cdot\frac{4(\partial_\sigma a(X_{t_i},\sigma_0))^2}{c(X_{t_i},\sigma_0)}1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}+\frac{1}{n}\sum_{i=0}^{n-1}R(\theta_0,\Delta_{n,i}^{\beta\wedge\delta_1\wedge(\beta-\frac{1}{4})},X_{t_i}).$$

The first term here above converges to $8\int_{\mathbb{R}}(\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2\pi(dx)$ as a consequence of the third point of Proposition 5 while the second one clearly goes to zero in norm 1 and so in probability thanks to the polynomial growth of R, its property (9) and the assumption we made $\beta>\frac{1}{4}$. It follows the first convergence of (56). To obtain the second one we observe it is

$$\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}\mathbb{E}_i[\tilde{\zeta}_i^{2+r}(\theta_0)]\leq\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}(\frac{-2\partial_\sigma m(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^{2+r}1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}\mathbb{E}_i[(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^{2+r}\varphi^{2+r}_{\Delta^\beta_{n,i}}(\Delta_i X)]+$$

$$+\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}(\frac{\partial_\sigma m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^{2+r}1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}(c+\frac{c}{m_2^{2+r}(\mu_0,\sigma_0,X_{t_i})}\mathbb{E}_i[(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^{2(2+r)}\varphi^{2(2+r)}_{\Delta^\beta_{n,i}}(\Delta_i X)])\leq$$

$$\leq\Delta_n^{(1+\frac{r}{2})\wedge(1+\beta(2+r))}\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i})+\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i})+\Delta_n^{0\wedge(1+2\beta(2+r)-(2+r))}\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i}),\tag{59}$$

where we have acted like before using the development (10) on m_2 and the second and the fourth point of Proposition 8. Besides, we have used the fourth point of Lemma 3 with $k=2+r$ and $k=2(2+r)$, respectively. It is now clear that the first two terms of (59) go to zero in norm 1 and so in probability for $n\to\infty$. Concerning the third one, if the minimum between 0 and $1+2\beta(2+r)-(2+r)$ is 0 it is exactly like the second one and so we know it goes to zero in probability, otherwise it can be seen as $\frac{1}{(n\Delta_n)^{\frac{r}{2}}}\Delta_n^{1+2\beta(2+r)-(2+r)+\frac{r}{2}}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i})$, that goes to zero since $n\Delta_n\to\infty$ for $n\to\infty$ and because of the fact that the exponent on Δ_n is always positive. Indeed $1+2\beta(2+r)-(2+r)+\frac{r}{2}>0$ iff $2\beta(2+r)>1+\frac{r}{2}$, that is $\beta>\frac{1+\frac{r}{2}}{2(2+r)}=\frac{1}{4}$.
To conclude, we prove the convergence (57). We have

$$\frac{1}{\sqrt{nT_n}}\sum_{i=0}^{n-1}|\mathbb{E}_i[\zeta_i(\theta_0)\tilde{\zeta}_i(\theta_0)]|=\frac{1}{\sqrt{nT_n}}\sum_{i=0}^{n-1}|\mathbb{E}_i\big[[(\frac{4\partial_\mu m\,\partial_\sigma m}{m_2^2})(\mu_0,\sigma_0,X_{t_i})(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2+$$

$$-2(\frac{\partial_\mu m\,\partial_\sigma m_2+\partial_\mu m_2\,\partial_\sigma m}{m_2^2})(\mu_0,\sigma_0,X_{t_i})(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))(1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})+\tag{60}$$

$$+\frac{\partial_\mu m_2\,\partial_\sigma m_2}{m_2^2})(\mu_0,\sigma_0,X_{t_i})(1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})^2]\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)\big]|1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}.$$

Now using the four points of Proposition 8, the first, second, third and fifth points of Lemma 3 and the lemma here above we get that (60) is upper bounded by

$$\frac{1}{n\sqrt{\Delta_n}}\sum_{i=0}^{n-1}R(\theta_0,\Delta_{n,i},X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{1}{3}+\beta},X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{4}{3}+\beta},X_{t_i})\leq\Delta_n^{\beta-\frac{1}{6}}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i}),$$

which converges to zero in norm 1 and so in probability since we have chosen $\beta>\frac{1}{4}>\frac{1}{6}$. □

8.3.2 Proof of Lemma 11.

Proof. Point 1.
We start showing the convergence of $C_n(\theta_0)$ to B. We observe that

$$\partial^2_{\mu\sigma}U_n(\mu_0,\sigma_0)=\sum_{i=0}^{n-1}[\frac{-2(X_{t_{i+1}}-m)\partial^2_{\mu\sigma}m}{m_2}-2\frac{\partial_\mu m\partial_\sigma m}{m_2}+\frac{2(X_{t_{i+1}}-m)\partial_\mu m\partial_\sigma m_2}{m_2^2}+\frac{2(X_{t_{i+1}}-m)\partial_\mu m_2\partial_\sigma m}{m_2^2}+\tag{61}$$

$$+\frac{2(X_{t_{i+1}}-m)^2\partial_\mu m_2\partial_\sigma m_2}{m_2^3}+\frac{\partial^2_{\mu\sigma}m_2}{m_2}(1-\frac{(X_{t_{i+1}}-m)^2}{m_2})-\frac{\partial_\mu m_2\partial_\sigma m_2}{m_2^2}]\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}=:\sum_{j=1}^{7}\sum_{i=0}^{n-1}I_{i,j}^n$$

where, for shortness, we omit that m, m_2 and their derivatives are calculated in (μ_0,σ_0,X_{t_i}).
In order to show that $\frac{1}{\sqrt{nT_n}}\partial^2_{\mu\sigma}U_n(\mu_0,\sigma_0)\xrightarrow{\mathbb{P}}0$ we will use repeatedly Lemma 9 in [14] and the estimation of the derivatives of m and m_2 gathered in Propositions 8 and 9.
$\sum_{i=0}^{n-1}I_{i,1}^n$ goes to zero in probability for $n\to\infty$ because $I_{i,1}^n$ is centered and, using also the first point of Lemma 3

$$\frac{1}{T_n}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[(I_{i,1}^n)^2]\leq\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,\Delta_{n,i},X_{t_i})R(\theta,\Delta_{n,i},X_{t_i})\leq\frac{1}{n\Delta_n}\frac{1}{n}\Delta_n^2\sum_{i=0}^{n-1}R(\theta,1,X_{t_i}),$$

which converges to zero in norm 1 and so in probability as a consequence of the polynomial growth of R and the third point of Lemma 2. $\sum_{i=0}^{n-1}I_{i,2}^n$ goes to zero in norm 1 and so in probability. Indeed, using the development (10) of m_2 and the first two point of Proposition 8, it is

$$\mathbb{E}[\frac{1}{\sqrt{nT_n}}|\sum_{i=0}^{n-1}I_{i,2}^n|]\leq\Delta_n^{-\frac{1}{2}}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|R(\theta,\Delta_n,X_{t_i})|]\leq c\Delta_n^{\frac{1}{2}},$$

which goes to zero.
Concerning $I_{i,3}^n$, it is still centered and from the first and fourth points of Proposition 8 and the first point of Lemma 3 it follows

$$\frac{1}{T_n}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[(I_{i,3}^n)^2]\leq\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,1,X_{t_i})R(\theta,\Delta_{n,i},X_{t_i})\leq\frac{1}{n\Delta_n}\frac{1}{n}\Delta_n\sum_{i=0}^{n-1}R(\theta,1,X_{t_i}),$$

which converges to 0 in norm 1 and so in probability. Hence, $\frac{1}{\sqrt{nT_n}}\sum_{i=0}^{n-1}I_{i,3}^n\xrightarrow{\mathbb{P}}0$ from Lemma 9 in [14].
The same applies to $I_{i,4}^n$, which squared is upper bounded by $\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,\Delta_{n,i}^2,X_{t_i})R(\theta,\Delta_{n,i},X_{t_i})\leq\frac{1}{n\Delta_n}\frac{1}{n}\Delta_n^3\sum_{i=0}^{n-1}R(\theta,1,X_{t_i})$.
On $\sum_{i=0}^{n-1}I_{i,5}^n$ we prove the convergence in norm 1 and so we have the convergence in probability: from the first point of Lemma 3, the development (10) of m_2 and the third and the fourth points of Proposition 8, it is

$$\frac{1}{\sqrt{nT_n}}\sum_{i=0}^{n-1}\mathbb{E}[|I_{i,5}^n|]\leq\frac{1}{n\sqrt{\Delta_n}}\sum_{i=0}^{n-1}R(\theta,\Delta_{n,i},X_{t_i})R(\theta,1,X_{t_i})\leq\Delta_n^{\frac{1}{2}}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,1,X_{t_i}),$$

which goes to zero.
We observe that, as a consequence of the definition of m_2, $I_{i,6}^n$ is centered. In order to apply Lemma 9 in [14] we evaluate its squared value, that is

$$\frac{1}{T_n}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[(I_{i,6}^n)^2]\leq\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,\Delta_{n,i},X_{t_i})R(\theta,1,X_{t_i})\leq\Delta_n\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,1,X_{t_i}),$$

where we have also used the estimation of the mixed second derivative of m_2 contained in Proposition 9.
To conclude the proof about the mixed derivative of the contrast function, we observe that $\sum_{i=0}^{n-1}I_{i,7}^n$ converges to 0 in L^1 from the third and the fourth point of Proposition 8 and the boundedness of φ. We obtain

$$\frac{1}{\sqrt{nT_n}}\mathbb{E}[|\sum_{i=0}^{n-1}I_{i,7}^n|]\leq\Delta_n^{-\frac{1}{2}}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta,\Delta_{n,i},X_{t_i})]\leq c\Delta_n^{\frac{1}{2}},$$

that goes to 0 as we wanted.
The next step is to prove that $\frac{1}{T_n}\partial^2_\mu U_n(\mu_0,\sigma_0)\xrightarrow{\mathbb{P}}-2\int_{\mathbb{R}}(\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2\pi(dx)$. In order to do it we compute $\partial^2_\mu U_n(\mu_0,\sigma_0)$ and we observe it is exactly like (61) but all the derivatives are with respect to μ. For such a reason we keep referring to (61) and we write $\partial^2_\mu U_n(\mu_0,\sigma_0)=:\sum_{j=1}^{7}\sum_{i=0}^{n-1}\tilde{I}_{i,j}^n$. We are going to show, in particular, that $\frac{1}{T_n}\sum_{i=0}^{n-1}\tilde{I}_{i,2}^n$ converges to the wanted integral, while $\frac{1}{T_n}(\sum_{i=0}^{n-1}\tilde{I}_{i,1}^n+\sum_{j=3}^{7}\sum_{i=0}^{n-1}\tilde{I}_{i,j}^n)\xrightarrow{\mathbb{P}}0$. Indeed, we observe that $\tilde{I}_{i,1}^n$, $\tilde{I}_{i,3}^n$, $\tilde{I}_{i,4}^n$, $\tilde{I}_{i,6}^n$ are still centered and, using Lemma 3 and Propositions 8 and 9 it is easy to show that their squared values are upper bounded in the following way:

$$\frac{1}{T_n^2}\sum_{i=0}^{n-1}\mathbb{E}_i[(\tilde{I}_{i,1}^n)^2]\leq\frac{1}{n\Delta_n}\Delta_n^{-1}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,\Delta_{n,i},X_{t_i})R(\theta,1,X_{t_i})\leq\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,1,X_{t_i}),$$

that goes to zero in norm 1 and so in probability since $n\Delta_n \to \infty$ for $n \to \infty$.

$$\frac{1}{T_n^2}\sum_{i=0}^{n-1}\mathbb{E}_i[(\tilde{I}^n_{i,3})^2] \leq \frac{1}{n\Delta_n}\Delta_n^{-1}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,\Delta_{n,i},X_{t_i})R(\theta,\Delta^2_{n,i},X_{t_i}) \leq \Delta_n^2\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,1,X_{t_i}) \xrightarrow{\mathbb{P}} 0.$$

Now $\tilde{I}^n_{i,4}$ and $\tilde{I}^n_{i,3}$ are exactly the same quantity and so the estimation here above clearly holds also for $\tilde{I}^n_{i,4}$ instead of $\tilde{I}^n_{i,3}$. Concerning $\tilde{I}^n_{i,6}$, we have

$$\frac{1}{T_n^2}\sum_{i=0}^{n-1}\mathbb{E}_i[(\tilde{I}^n_{i,6})^2] \leq \frac{1}{n\Delta_n}\Delta_n^{-1}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,\Delta^2_{n,i},X_{t_i})R(\theta,1,X_{t_i}) \leq \Delta_n\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,1,X_{t_i}) \xrightarrow{\mathbb{P}} 0.$$

The application of Lemma 9 in [14] gives us $\frac{1}{T_n}\sum_{i=0}^{n-1}(\tilde{I}^n_{i,1}+\tilde{I}^n_{i,3}+\tilde{I}^n_{i,4}+\tilde{I}^n_{i,6}) \xrightarrow{\mathbb{P}} 0$. We now prove the convergence to 0 in norm 1 of $\frac{1}{T_n}\sum_{i=0}^{n-1}(\tilde{I}^n_{i,5}+\tilde{I}^n_{i,7})$. Indeed, using again the first point of Lemma 3, the development (10) of m_2 and the last two points of Proposition 9 it is $\frac{1}{T_n}\sum_{i=0}^{n-1}\mathbb{E}[|\tilde{I}^n_{i,5}+\tilde{I}^n_{i,7}|] \leq$

$$\leq \frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta,\Delta_{n,i},X_{t_i})R(\theta,\Delta_{n,i},X_{t_i})+R(\theta,\Delta^2_{n,i},X_{t_i})] \leq \Delta_n\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta,1,X_{t_i})] \leq c\Delta_n,$$

which clearly goes to 0.
Concerning the principal term $\frac{1}{T_n}\sum_{i=0}^{n-1}\tilde{I}^n_{i,2}$, we observe that using (34) and the first point of Proposition 8, it is

$$\frac{(\partial_\mu m)^2}{m_2}(\mu_0,\sigma_0,X_{t_i}) = \frac{(\Delta_{n,i}\partial_\mu b(\mu_0,X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{5}{2}-\beta-\epsilon},X_{t_i}))^2}{\Delta_{n,i}c(\sigma_0,X_{t_i})}(1-\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma_0)}-\Delta_{n,i}r(\mu_0,\sigma_0,X_{t_i})+$$

$$+R(\theta_0,\Delta_{n,i}^{\bar{r}},X_{t_i})) = \frac{\Delta_{n,i}(\partial_\mu b(\mu_0,X_{t_i}))^2}{c(\sigma_0,X_{t_i})}+R(\theta_0,\Delta_{n,i}^{(\frac{5}{2}-\beta-\epsilon)\wedge(1+\delta_1)},X_{t_i}),$$

with $\bar{r}=2\wedge(1+\delta_2)\wedge 2\delta_1$, as defined below (34). In the last equality we have used that the other terms are negligible. Now we have that

$$\frac{1}{T_n}\sum_{i=0}^{n-1}-2\frac{\Delta_{n,i}(\partial_\mu b(\mu_0,X_{t_i}))^2}{c(\sigma_0,X_{t_i})}\varphi_{\Delta^\beta_{n,i}}(\Delta X_i)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}} \xrightarrow{\mathbb{P}} -2\int_{\mathbb{R}}(\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2\pi(dx),$$

as a consequence of the second point of Proposition 5 while
$\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}-2R(\theta_0,\Delta_{n,i}^{(\frac{5}{2}-\beta-\epsilon)\wedge(1+\delta_1)},X_{t_i})\varphi_{\Delta^\beta_{n,i}}(\Delta X_i)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}$ is upper bounded in norm 1 by

$$\Delta_n^{(\frac{3}{2}-\beta-\epsilon)\wedge\delta_1}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta_0,1,X_{t_i})] \leq c\Delta_n^{(\frac{3}{2}-\beta-\epsilon)\wedge\delta_1},$$

that converges to 0 since the exponent on Δ_n is always positive.
To prove the first point of Lemma 11 we are left to show the convergence of $\frac{1}{n}\partial^2_\sigma U_n(\mu_0,\sigma_0)$.
Again, we still refer to (61) observing that that the only difference is that all the derivatives are with respect to σ. We write $\partial^2_\sigma U_n(\mu_0,\sigma_0) =: \sum_{j=1}^7\sum_{i=0}^{n-1}\hat{I}^n_{i,j}$.
We keep using Lemma 9 in [14] joint with the development (10) and Propositions 8 and 9 to show that the centered terms go to zero in probability, that is

$$\frac{1}{n}\sum_{i=0}^{n-1}(\hat{I}^n_{i,1}+\hat{I}^n_{i,3}+\hat{I}^n_{i,4}+\hat{I}^n_{i,6})\varphi_{\Delta^\beta_{n,i}}(\Delta X_i)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}} \xrightarrow{\mathbb{P}} 0.$$

Moreover,

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|\hat{I}^n_{i,2}|] \leq \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta,\Delta_{n,i},X_{t_i})] \leq c\Delta_n \to 0.$$

We are left to deal with the principal terms $\hat{I}^n_{i,5}$ and $\hat{I}^n_{i,7}$ and so we study the convergence of
$\frac{1}{n}\sum_{i=0}^{n-1}\frac{(\partial_\sigma m_2)^2}{m_2^2}(\frac{2(X_{t_{i+1}}-m)^2}{m_2}-1)\varphi_{\Delta^\beta_{n,i}}(\Delta X_i)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}$. From the development (10) of m_2 and the development of $\partial_\sigma m_2$ stated in the fourth point of Proposition 8 it follows that the conditional expected value of the quantity here above is
$\frac{1}{n}\sum_{i=0}^{n-1}\frac{(2\Delta_{n,i}\partial_\sigma a(X_{t_i},\sigma_0)a(X_{t_i},\sigma_0))^2}{(\Delta_{n,i}a^2(X_{t_i},\sigma_0))^2}(\frac{2(X_{t_{i+1}}-m)^2}{\Delta_{n,i}a^2(X_{t_i},\sigma_0)}-1)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}$ plus a negligible term that comes from

the developments (10) and (34) and it converges to zero in norm 1 and so in probability.
The principal term is therefore such that, using the second point of Proposition 6 and the third of Proposition 5, we get

$$\frac{1}{n}\sum_{i=0}^{n-1}(\hat{I}^n_{i,5}+\hat{I}^n_{i,7}) \xrightarrow{\mathbb{P}} 4\int_{\mathbb{R}}(\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2\pi(dx).$$

Point 2.
We start proving that $\frac{1}{\sqrt{nT_n}}\sup_{|\tilde{\theta}|\leq\epsilon_n}|\partial^2_{\mu\sigma}U_n(\theta_0+\tilde{\theta})-\partial^2_{\mu\sigma}U_n(\theta_0)|$ goes to 0 in probability for ϵ_n that goes to 0.
In order to do that, it is enough to show that the sequence $\frac{1}{\sqrt{nT_n}}\partial^2_{\mu\sigma}U_n(\theta)$ is tight, which is implied by $\sup_n\frac{1}{\sqrt{nT_n}}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta(\partial^2_{\mu\sigma}U_n(\mu,\sigma))|]<\infty$ (see Corollary B.1 in [33]), for $\vartheta=\mu$ or $\vartheta=\sigma$.
We observe it is

$$\partial^3_{\mu\sigma\vartheta}U_n(\mu,\sigma):=\sum_{i=0}^{n-1}[\frac{2\partial_\vartheta m\partial^2_{\mu\sigma}m-2(X_{t_{i+1}}-m)\partial^3_{\mu\sigma\vartheta}m}{m_2}+\frac{2(X_{t_{i+1}}-m)\partial^2_{\mu\sigma}m\partial_\vartheta m_2}{m_2^2}-2\frac{(\partial^2_{\mu\vartheta}m\partial_\sigma m+\partial_\mu m\partial^2_{\sigma\vartheta}m)}{m_2}+$$

$$+\frac{2\partial_\mu m\partial_\sigma m\partial_\vartheta m_2}{m_2^2}-\frac{2\partial_\vartheta m\partial_\mu m\partial_\sigma m_2}{m_2^2}+\frac{2(X_{t_{i+1}}-m)(\partial^2_{\mu\vartheta}m\partial_\sigma m_2+\partial_\mu m\partial^2_{\sigma\vartheta}m_2)}{m_2^2}-\frac{4(X_{t_{i+1}}-m)\partial_\mu m\partial_\sigma m_2\partial_\vartheta m_2}{m_2^3}+$$

$$-\frac{2\partial_\vartheta m\partial_\mu m_2\partial_\sigma m}{m_2^2}+\frac{2(X_{t_{i+1}}-m)(\partial^2_{\mu\vartheta}m_2\partial_\sigma m+\partial_\mu m_2\partial^2_{\sigma\vartheta}m)}{m_2^2}-\frac{4(X_{t_{i+1}}-m)\partial_\mu m_2\partial_\sigma m\partial_\vartheta m_2}{m_2^3}+$$

$$-\frac{4(X_{t_{i+1}}-m)\partial_\vartheta m\partial_\mu m_2\partial_\sigma m_2}{m_2^3}+\frac{2(X_{t_{i+1}}-m)^2(\partial^2_{\mu\vartheta}m_2\partial_\sigma m_2+\partial_\mu m_2\partial^2_{\sigma\vartheta}m_2)}{m_2^3}-\frac{6(X_{t_{i+1}}-m)^2\partial_\mu m_2\partial_\sigma m_2\partial_\vartheta m_2}{m_2^4}+$$

$$+(\frac{\partial^3_{\mu\sigma\vartheta}m_2}{m_2}-\frac{\partial^2_{\mu\sigma}m_2\partial_\vartheta m_2}{m_2^2})(1-\frac{(X_{t_{i+1}}-m)^2}{m_2})+\frac{\partial^2_{\mu\sigma}m_2}{m_2}(\frac{2(X_{t_{i+1}}-m)\partial_\vartheta m}{m_2}+\frac{(X_{t_{i+1}}-m)^2\partial_\vartheta m_2}{m_2^2})+$$

$$-\frac{\partial^2_{\mu\vartheta}m_2\partial_\sigma m_2+\partial_\mu m_2\partial^2_{\sigma\vartheta}m_2}{m_2^2}+\frac{2\partial_\mu m_2\partial_\sigma m_2\partial_\vartheta m_2}{m_2^3}]\varphi_{\Delta^\beta_{n,i}}(\Delta X_i)1_{\{|X_{t_i}|\leq\Delta^{-k}_{n,i}\}}=:\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\mu,\sigma,\vartheta).$$

Now using Assumption *Ad*, the estimation on the derivatives of m and m_2 gathered in Propositions 8, 9 and 10 and the inequalities (40), (41) and (42) it follows

$$\mathbb{E}[\sup_{\mu,\sigma}|\partial_\mu(\partial^2_{\mu\sigma}U_n(\mu,\sigma))|]=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\mu,\sigma,\mu)|]\leq cn\Delta_n^{\frac{1}{2}}.$$

In the same way we get

$$\mathbb{E}[\sup_{\mu,\sigma}|\partial_\sigma(\partial^2_{\mu\sigma}U_n(\mu,\sigma)|]=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\mu,\sigma,\sigma)|]\leq cn\Delta_{n,i}^{\frac{1}{2}}.$$

Hence, for both $\vartheta=\mu$ and $\vartheta=\sigma$ we can say it is $\sup_n\frac{1}{\sqrt{nT_n}}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta(\partial^2_{\mu\sigma}U_n(\mu,\sigma)|]\leq c<\infty$; that implies the tightness of our sequence and so that $\frac{1}{\sqrt{nT_n}}\sup_{|\tilde{\theta}|\leq\epsilon_n}|\partial^2_{\mu\sigma}U_n(\theta_0+\tilde{\theta})-\partial^2_{\mu\sigma}U_n(\theta_0)|$ goes to 0 in probability for ϵ_n that goes to 0.
To prove the convergence to 0 in probability of $\frac{1}{n}\sup_{|\tilde{\theta}|\leq\epsilon_n}|\partial^2_\sigma U_n(\theta_0+\tilde{\theta})-\partial^2_\sigma U_n(\theta_0)|$ for ϵ_n that goes to 0 we act in the same way: we show that the sequence $\frac{1}{n}\partial^2_\sigma U_n(\theta)$ is tight through the criterion $\sup_n\frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta(\partial^2_\sigma U_n(\mu,\sigma))|]<\infty$.
We observe that, computing the derivative with respect to ϑ of $\partial^2_\sigma U_n(\mu,\sigma)$, we obtain 17 terms analogous to the case just studied, with the only difference that also the derivatives that were with respect to μ are now with respect to σ. In particular, it is $\partial_\vartheta(\partial^2_\sigma U_n(\mu,\sigma))=\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\sigma,\sigma,\vartheta)$.
We still use Assumption *Ad*, the estimation on the derivatives of m and m_2 gathered in Propositions 8, 9 and 10 and the inequalities (40), (41) and (42) to prove that $\mathbb{E}[\sup_{\mu,\sigma}|\partial_\sigma(\partial^2_\sigma U_n(\mu,\sigma)|]<\infty$. Indeed, it is

$$\mathbb{E}[\sup_{\mu,\sigma}|\partial_\sigma(\partial^2_\sigma U_n(\mu,\sigma)|]=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\sigma,\sigma,\sigma)|]\leq c.$$

Moreover, since the order in which we compute the derivatives of the contrast function commute, we have

$$\mathbb{E}[\sup_{\mu,\sigma}|\partial_\mu(\partial^2_\sigma U_n(\mu,\sigma)|]=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\sigma,\sigma,\mu)|]=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\mu,\sigma,\sigma)|]\leq\sum_{i=0}^{n-1}c\Delta_{n,i}^{\frac{1}{2}}.$$

We can therefore say that, for both $\vartheta = \mu$ and $\vartheta = \sigma$, it is $\sup_n \frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma} |\partial_\vartheta(\partial^2_\sigma U_n(\mu,\sigma)|] \leq c < \infty$; that implies the tightness.
We are now left to show that $\frac{1}{T_n}\sup_{|\tilde{\theta}|\leq\epsilon_n} |\partial^2_\mu U_n(\theta_0+\tilde{\theta}) - \partial^2_\mu U_n(\theta_0)| \xrightarrow{\mathbb{P}} 0$ for $\epsilon_n \to 0$. We still consider the notation introduced in the first point for which $\partial^2_\mu U_n(\theta) =: \sum_{j=1}^{7}\sum_{i=0}^{n-1} \tilde{I}^n_{i,j}(\theta)$ with $\partial^2_\mu U_n(\theta)$ that is as in (61) but both the derivatives are calculated with respect to μ.
From Proposition 7 we already know that $\frac{1}{T_n}\sum_{i=0}^{n-1} \tilde{I}^n_{i,j}(\theta)$ are tight sequences for $j \in \{1,3,4\}$; having taken as $g_{i,n}(\theta, X_{t_i})$ respectively $\frac{-2\partial^2_\mu m(\mu,\sigma,X_{t_i})}{m_2(\mu,\sigma,X_{t_i})}$ and $\frac{2\partial_\mu m(\mu,\sigma,X_{t_i})\partial_\mu m_2(\mu,\sigma,X_{t_i})}{m_2^2(\mu,\sigma,X_{t_i})}$, twice. We see that the assumptions required on $g_{i,n}$ hold as a consequence of the estimation on the derivatives of m and m_2 gathered in Propositions 8 and 9 and the development Ad of m_2.
We also show the tightness of the other terms proving that, for both $\vartheta = \mu$ and $\vartheta = \sigma$,
$\sup_n \frac{1}{T_n}\mathbb{E}[\sup_{\mu,\sigma} |\sum_{i=0}^{n-1} \partial_\vartheta(\tilde{I}^n_{i,2}+\tilde{I}^n_{i,5}+\tilde{I}^n_{i,6}+\tilde{I}^n_{i,7})|] \leq c$. Indeed, using the estimation on the first derivatives of m and m_2 gathered in Proposition 8 and the development Ad of m_2 it is

$$\sup_n \frac{1}{T_n}\mathbb{E}[\sup_{\mu,\sigma} |\sum_{i=0}^{n-1} \partial_\mu \tilde{I}^n_{i,2}|] \leq \sup_n \frac{c}{n\Delta_n}\sum_{i=0}^{n-1}(\Delta_{n,i} + \Delta^2_{n,i}) \leq c,$$

$$\sup_n \frac{1}{T_n}\mathbb{E}[\sup_{\mu,\sigma} |\sum_{i=0}^{n-1} \partial_\sigma \tilde{I}^n_{i,2}|] \leq \sup_n \frac{c}{n\Delta_n}\sum_{i=0}^{n-1}(\Delta_{n,i} + \Delta_{n,i}) \leq c.$$

In the same way, using Assumption Ad, the estimation on the derivatives of m and m_2 gathered in Propositions 8, 9 and 10 and the inequalities (40), (41) and (42) it follows

$$\sup_n \frac{1}{T_n}\mathbb{E}[\sup_{\mu,\sigma} |\sum_{i=0}^{n-1} \partial_\mu(\tilde{I}^n_{i,5} + \tilde{I}^n_{i,6} + \tilde{I}^n_{i,7})|] \leq \sup_n \frac{c}{n\Delta_n}(\Delta_n^{\frac{5}{2}\wedge 2\wedge 3} + \Delta_n^{1\wedge 2\wedge\frac{3}{2}\wedge 2} + \Delta_n^{2\wedge 3}) \leq c;$$

$$\sup_n \frac{1}{n\Delta_n}\mathbb{E}[\sup_{\mu,\sigma} |\sum_{i=0}^{n-1} \partial_\sigma(\tilde{I}^n_{i,5} + \tilde{I}^n_{i,6} + \tilde{I}^n_{i,7})|] \leq \sup_n \frac{c}{n\Delta_n}(\Delta_n^{\frac{5}{2}\wedge 2\wedge 2} + \Delta_n^{1\wedge 1\wedge\frac{3}{2}\wedge 2} + \Delta_n^{1\wedge 2}) \leq c.$$

We have therefore proved that the sequence $\frac{1}{T_n}\partial^2_\mu U_n(\theta)$ is tight, which implies the convergence to zero in probability of $\frac{1}{T_n}\sup_{|\tilde{\theta}|\leq\epsilon_n} |\partial^2_\mu U_n(\theta_0+\tilde{\theta}) - \partial^2_\mu U_n(\theta_0)|$. □

8.3.3 Proof of Theorem 2.

Proof. By (54) we get

$$(\int_0^1 [C_n(\theta_0 + u(\hat{\theta}_n - \theta_0)) - C_n(\theta_0)]du + C_n(\theta_0))S_n = L_n(\theta_0).$$

We find that the matrix

$$\int_0^1 [C_n(\theta_0 + u(\hat{\theta}_n - \theta_0)) - C_n(\theta_0)]du + C_n(\theta_0) \tag{62}$$

converges in probability to the nonsingular matrix B. Hence, taking the limit on both sides after multiplying by the inverse of (62), we see by the continuous mapping theorem that $S_n \xrightarrow{d} B^{-1}L \sim N(0, K^{-1})$. The asymptotic normality of S_n is therefore proved. □

8.4 Proof of Proposition 1

We observe that, having assumed Ad on $\tilde{m}_2$ and as a consequence of the first point of A_ρ, the developments (10) and (31) keep holding true with $\tilde{m}$ and $\tilde{m}_2$ instead of m and m_2 with the only difference that the function $r_1(\mu,\sigma,x)$ in the development of $\tilde{m}$ contains also the rest function $R(\theta, \Delta^{\rho_1}_{n,i}, x)$. Now it is possible to prove on $\tilde{m}$ and $\tilde{m}_2$ every result stated in Section 8.1 following the proof we give with $\tilde{m}$ and $\tilde{m}_2$ replacing m and m_2, since the only tools we use are the here above discussed developments. Moreover, the results stated in Section 8.2 hold true also on $\tilde{m}$ and $\tilde{m}_2$ for the third point of A_ρ.
The substantial difference between m and m_2 and their approximation is that we have defined m and m_2 as in (5) and (6) on purpose to make $L_n(\theta_0)$ a triangular array of martingale increments without requiring any constraint on the rate at which the step discretization has to go to zero. Defining $\tilde{U}_n(\mu,\sigma)$ the contrast function in which we have replaced m and m_2 with their approximation $\tilde{m}$ and $\tilde{m}_2$, we have that

$$\tilde{L}_n(\theta_0) := \begin{pmatrix} -\frac{1}{\sqrt{T_n}}\partial_\mu \tilde{U}_n(\mu_0,\sigma_0) \\ -\frac{1}{\sqrt{n}}\partial_\sigma \tilde{U}_n(\mu_0,\sigma_0) \end{pmatrix}$$

is no longer a triangular array of martingale increments regardless and so we have to provide an alternative to Lemma 10, which is gathered in the following lemma.

Lemma 13. *Suppose that Assumptions A1-A8, Ad, A_{Step} and Aρ hold, with $0 < k < k_0$, and that $\sqrt{n}\Delta_n^{\rho_1 - 1/2} \to 0$ and $\sqrt{n}\Delta_n^{\rho_2 - 1} \to 0$ as $n \to \infty$. Then, as $n \to \infty$,*

$$\tilde{L}_n(\theta_0) \xrightarrow{d} L \sim N(0, K'),$$

where $K' = \begin{pmatrix} 4\int_{\mathbb{R}} (\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2 \pi(dx) & 0 \\ 0 & 8\int_{\mathbb{R}} (\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2 \pi(dx) \end{pmatrix}$.

Proof. The result follows from a combination of Theorem 3.2 and Theorem 3.4 in [18] (c.f. also Section A.2 in the Appendix of [33]). We get the lemma proven if, for a constant $r > 0$, the following convergences hold.
We define ζ_i and $\tilde{\zeta}_i$ such that $\partial_\mu \tilde{U}_n(\mu_0, \sigma_0) =: \sum_{i=0}^{n-1} \zeta_i(\theta_0)$ and $\partial_\sigma \tilde{U}_n(\mu_0, \sigma_0) =: \sum_{i=0}^{n-1} \tilde{\zeta}_i(\theta_0)$. Then it must be

$$\frac{1}{\sqrt{T_n}} \sum_{i=0}^{n-1} |\mathbb{E}_i[\zeta_i(\theta_0)]| \xrightarrow{\mathbb{P}} 0 \qquad \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} |\mathbb{E}_i[\tilde{\zeta}_i(\theta_0)]| \xrightarrow{\mathbb{P}} 0, \tag{63}$$

$$\frac{1}{T_n} \sum_{i=0}^{n-1} \mathbb{E}_i[\zeta_i^2(\theta_0)] \xrightarrow{\mathbb{P}} 4\int_{\mathbb{R}} (\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2 \pi(dx) \qquad \frac{1}{(\sqrt{T_n})^{2+r}} \sum_{i=0}^{n-1} \mathbb{E}_i[\zeta_i^{2+r}(\theta_0)] \xrightarrow{\mathbb{P}} 0, \tag{64}$$

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}_i[\tilde{\zeta}_i^2(\theta_0)] \xrightarrow{\mathbb{P}} 8\int_{\mathbb{R}} (\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2 \pi(dx) \qquad \frac{1}{(\sqrt{n})^{2+r}} \sum_{i=0}^{n-1} \mathbb{E}_i[\tilde{\zeta}_i^{2+r}(\theta_0)] \xrightarrow{\mathbb{P}} 0, \tag{65}$$

$$\frac{1}{\sqrt{nT_n}} \sum_{i=0}^{n-1} |\mathbb{E}_i[\zeta_i(\theta_0)\tilde{\zeta}_i(\theta_0)]| \xrightarrow{\mathbb{P}} 0. \tag{66}$$

It is enough to follow the proof of (55), (56) and (57) with $\tilde{m}$ and $\tilde{m}_2$ instead of m and m_2 to get (64), (65) and (66).
We are left to show (63). We observe that, by the definition of m and the first point of A_ρ, we have

$$|\mathbb{E}_i[(X_{t_{i+1}} - \tilde{m}(\mu_0, \sigma_0, X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)]| = |\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] +$$

$$+\mathbb{E}_i[(m(\mu_0, \sigma_0, X_{t_i}) - \tilde{m}(\mu_0, \sigma_0, X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)]| \leq R(\theta_0, \Delta_{n,i}^{\rho_1}, X_{t_i}). \tag{67}$$

In the same way, using also the definition of m_2 and the estimation which assesses the quality of the approximation of m_2 through $\tilde{m}_2$, still in the first point of A_ρ, we get

$$\mathbb{E}_i[(1 - \frac{(X_{t_{i+1}} - \tilde{m}(\mu_0, \sigma_0, X_{t_i}))^2}{\tilde{m}_2(\mu_0, \sigma_0, X_{t_i})})\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] =$$

$$= \mathbb{E}_i[(1 - \frac{(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))^2}{m_2(\mu_0, \sigma_0, X_{t_i})})\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] +$$

$$+\mathbb{E}_i[(\frac{(X_{t_{i+1}} - \tilde{m}(\mu_0, \sigma_0, X_{t_i}))^2}{\tilde{m}_2(\mu_0, \sigma_0, X_{t_i})} - \frac{(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))^2}{m_2(\mu_0, \sigma_0, X_{t_i})})\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] =$$

$$= 0 + (\frac{1}{\tilde{m}_2(\mu_0, \sigma_0, X_{t_i})} - \frac{1}{m_2(\mu_0, \sigma_0, X_{t_i})})\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))^2\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)] +$$

$$+\mathbb{E}_i[\frac{(m(\mu_0, \sigma_0, X_{t_i}) - \tilde{m}(\mu_0, \sigma_0, X_{t_i}))^2 + 2(m(\mu_0, \sigma_0, X_{t_i}) - \tilde{m}(\mu_0, \sigma_0, X_{t_i}))(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))}{\tilde{m}_2(\mu_0, \sigma_0, X_{t_i})}\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)].$$

Now, using also the development A_d for m_2 and $\tilde{m}_2$ and the first and the third point of Lemma 3, the absolute value of the equation here above is upper bounded by

$$R(\theta_0, \Delta_{n,i}^{\rho_2 - 1}, X_{t_i}) + R(\theta_0, \Delta_{n,i}^{2\rho_1 - 1}, X_{t_i}) + R(\theta_0, \Delta_{n,i}^{\rho_1}, X_{t_i}) = R(\theta_0, \Delta_{n,i}^{(\rho_2 - 1)\wedge\rho_1}, X_{t_i}). \tag{68}$$

We compute hereafter the derivatives of $\tilde{U}_n$ with respect to μ, obtaining

$$\mathbb{E}_i[\zeta_i(\theta_0)] = \frac{-2\partial_\mu \tilde{m}(\mu_0, \sigma_0, X_{t_i})}{\tilde{m}_2(\mu_0, \sigma_0, X_{t_i})}\mathbb{E}_i[(X_{t_{i+1}} - \tilde{m}(\mu_0, \sigma_0, X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}} + \tag{69}$$

$$+\frac{\partial_\mu \tilde{m}_2(\mu_0, \sigma_0, X_{t_i})}{\tilde{m}_2(\mu_0, \sigma_0, X_{t_i})}\mathbb{E}_i[(1 - \frac{(X_{t_{i+1}} - \tilde{m}(\mu_0, \sigma_0, X_{t_i}))^2}{m_2(\mu_0, \sigma_0, X_{t_i})})\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)]1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}}.$$

As a consequence of the third point of A_ρ, Proposition 8 still holds replacing the derivatives of m and m_2 with the derivatives of $\tilde{m}$ and $\tilde{m}_2$. Therefore, using also Ad and the equations (67) and (68), we have

$$\frac{1}{\sqrt{T_n}}\sum_{i=0}^{n-1}|\mathbb{E}_i[\zeta_i(\theta_0)]| \leq \frac{c}{\sqrt{n\Delta_n}}\sum_{i=0}^{n-1}(R(\theta_0,\Delta_{n,i}^{\rho_1},X_{t_i})+R(\theta_0,\Delta_{n,i}^{\rho_2\wedge\rho_1+1},X_{t_i})) \leq \frac{\Delta_n^{(\rho_1-\frac{1}{2})\wedge(\rho_2-\frac{1}{2})}}{\sqrt{n}}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i}),$$

which converges to 0 in norm 1 and so in probability for $\sqrt{n}\Delta_n^{(\rho_1-\frac{1}{2})\wedge(\rho_2-\frac{1}{2})} \to 0$ for $n\to\infty$, as we have required.
To show Lemma 13 holds true we have to prove finally that $\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}|\mathbb{E}_i[\tilde{\zeta}_i(\theta_0)]| \xrightarrow{\mathbb{P}} 0$ for $n\to\infty$, recalling that $\mathbb{E}_i[\tilde{\zeta}_i(\theta_0)]$ is defined as (69) but with the derivatives with respect to σ which take now the place of the derivatives with respect to μ. From Proposition 8, Ad and the equations (67) and (68), it follows

$$\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathbb{E}_i[\tilde{\zeta}_i(\theta_0)] \leq \frac{c}{\sqrt{n}}\sum_{i=0}^{n-1}(R(\theta_0,\Delta_{n,i}^{\rho_1},X_{t_i})+R(\theta_0,\Delta_{n,i}^{\rho_2-1\wedge\rho_1},X_{t_i})) \leq \frac{\Delta_n^{\rho_1\wedge(\rho_2-1)}}{\sqrt{n}}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i}),$$

which converges to 0 in norm 1 and so in probability for $\sqrt{n}\Delta_n^{\rho_1\wedge(\rho_2-1)} \to 0$ for $n\to\infty$, as we have required.
We get $\tilde{L}_n(\theta_0)$ is asymptotically normal, as we wanted. □

After having replaced Lemma 10 with the Lemma 13 just showed, it is enough to follow the proof of the asymptotic normality of the estimator $\hat{\theta_n}$ given in Sections 8.1−8.3 to get the asymptotic normality of $\tilde{\theta}_n$.

A Appendix

In this section we prove all the technical results we have introduced, starting from the preliminary results stated in Section 7.

A.1 Proof of limit theorems

We first show Proposition 5, observing that its last two points are the discretized version of the first point of Lemma 2.

A.1.1 Proof of Proposition 5

Proof. The first two points have already been proved in Proposition 3 of [2].
We want to show that $\frac{1}{n}\sum_{i=0}^{n-1}f(X_{t_i},\theta)$ converges in L^2 to $\int_{\mathbb{R}}f(x,\theta)\pi(dx)$.
Since $Var(\frac{1}{n}\sum_{i=0}^{n-1}f(X_{t_i},\theta)) \leq \frac{1}{n^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}Cov(f(X_{t_i},\theta),f(X_{t_j},\theta))$, we need to estimate the covariance.
We know that, under our assumptions, the process X is β- mixing with exponential decay (see [27]) that is $\exists\gamma>0$ such that $\beta_X(k)=O(e^{-\gamma k})$; with $\beta_X(k)$ as defined in Section 1.3.2 of [10]. If a process is β-mixing, then it is also α-mixing and so the following estimation holds (see Theorem 3 in Section 1.2.2 of [10])

$$|Cov(X_{t_i},X_{t_j})| \leq c\left\|X_{t_i}\right\|_p\left\|X_{t_j}\right\|_q\alpha^{\frac{1}{r}}(X_{t_i},X_{t_j})$$

with p, q and r such that $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1$. Using that $\alpha(X_{t_i},X_{t_j}) \leq \beta_X(|t_i-t_j|)=O(e^{-\gamma|t_i-t_j|})$, in our case the inequality here above becomes $|Cov(f(X_{t_i},\theta),f(X_{t_j},\theta))| \leq ce^{-\frac{1}{r}\gamma|t_i-t_j|}$, where we have also used the polynomial growth of f and the third point of Lemma 2 to include the two norms in the constant c.
We introduce a partition of $(0,T_n]$ based on the sets $A_k := (k\frac{T_n}{n},(k+1)\frac{T_n}{n}]$, for which $(0,T_n]=\cup_{k=0}^{n-1}A_k$. Now each point t_i in $(0,T_n]$ can be seen as $t_{k,h}$, where k identifies the particular set A_k to which the point belongs while, defining M_k as $|A_k|$, h is a number in $\{1,...,M_k\}$ which enumerates the points in each set. It follows

$$\frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}e^{-\frac{1}{r}\gamma|t_i-t_j|} \leq \frac{c}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\sum_{h_1=1}^{M_{k_1}}\sum_{h_2=1}^{M_{k_2}}e^{-\frac{1}{r}\gamma|t_{k_1,h_1}-t_{k_2,h_2}|} \leq \frac{ce^{\frac{1}{r}\frac{T_n}{n}}}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\sum_{h_1=1}^{M_{k_1}}\sum_{h_2=1}^{M_{k_2}}e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}},$$

where the last inequality is a consequence of the following estimation: for each $k_1,k_2\in\{0,...,n-1\}$ it is $|t_{k_1,h_1}-t_{k_2,h_2}| \geq |k_1-k_2|\frac{T_n}{n}-\frac{T_n}{n}$.
Now we observe that the exponent does not depend on h anymore, hence the last term here above can be upper bounded by $\frac{ce^{\frac{1}{r}\frac{T_n}{n}}}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}M_{k_1}M_{k_2}e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}}$.

Moreover, remarking that the length of each interval A_k is $\frac{T_n}{n}$, it is easy to show that we can always upper bound M_k with $\frac{T_n}{n}\frac{1}{\Delta_{min}}$, with $T_n = \sum_{i=0}^{n-1}\Delta_{n,i} \leq n\Delta_n$ and so $M_k \leq \frac{\Delta_n}{\Delta_{min}}$, that we have assumed bounded by a constant c_1.
Furthermore, still using that $T_n \leq n\Delta_n$, we have $e^{\frac{1}{r}\frac{T_n}{n}} \leq e^{\frac{1}{r}\Delta_n} \leq c$ because, by our hypothesis, Δ_{max} goes to 0 for $n \to \infty$. To conclude, we have to show that $\frac{c}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}} \to 0$ for $n \to \infty$. We define $j := k_1 - k_2$ and we apply a change of variable, getting

$$\frac{c}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}} \leq \frac{c}{n^2}\sum_{j=-(n-1)}^{n-1} e^{-\frac{1}{r}\gamma|j|\frac{T_n}{n}}|n-j| \leq \frac{c}{n}\sum_{j=-(n-1)}^{n-1} e^{-\frac{1}{r}\gamma|j|\Delta_{min}} \leq \frac{c}{n(1-e^{-\frac{1}{r}\gamma\Delta_{min}})} \leq \frac{c}{T_n},$$

that goes to 0 for n that goes to ∞. We therefore get $|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta) - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0$.
In order to show the third point we observe that

$$|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)1_{\left\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\right\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \leq |\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)1_{\left\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\right\}} - \frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)| +$$

$$+|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta) - \int_{\mathbb{R}} f(x,\theta)\pi(dx)|.$$

We have already proved that the second goes to 0 in probability, while the first is $|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)1_{\left\{|X_{t_i}|>\Delta_{n,i}^{-k}\right\}}|$, that converges to 0 in L^1 as a consequence of the polynomial growth of f and the third point of Lemma 2 and so in probability.
We act in the same way in order to show the fourth point, observing that, by the definition of φ, it is

$$|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)1_{\left\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\right\}}(\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i) - 1)| \leq \frac{c}{n}\sum_{i=0}^{n-1}|f(X_{t_i},\theta)|1_{\left\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\right\}}1_{\left\{|X_{t_i}|\geq\Delta_{n,i}^{\beta}\right\}}. \tag{70}$$

We observe that, since $\Delta X_i^c = \Delta X_i - \Delta X_i^J$, if $|\Delta X_i| \geq \Delta_{n,i}^{\beta}$ and $|\Delta X_i^J| < \frac{\Delta_{n,i}^{\beta}}{2}$, then $|\Delta X_i^c|$ must be more than $\frac{\Delta_{n,i}^{\beta}}{2}$. Hence

$$\mathbb{E}[1_{\left\{|\Delta X_i|\geq\Delta_{n,i}^{\beta}\right\}}] = \mathbb{E}[1_{\left\{|\Delta X_i|\geq\Delta_{n,i}^{\beta},|\Delta X_i^J|<\frac{\Delta_{n,i}^{\beta}}{2}\right\}}] + \mathbb{E}[1_{\left\{|\Delta X_i|\geq\Delta_{n,i}^{\beta},|\Delta X_i^J|\geq\frac{\Delta_{n,i}^{\beta}}{2}\right\}}] \leq$$

$$\leq \mathbb{P}(|\Delta X_i^c| \geq \frac{\Delta_{n,i}^{\beta}}{2}) + \mathbb{P}(|\Delta X_i^J| \geq \frac{\Delta_{n,i}^{\beta}}{2}) \leq c\frac{\mathbb{E}[|\Delta X_i^c|^r]}{\Delta_{n,i}^{\beta r}} + R(\theta,\Delta_{n,i},X_{t_i}) \leq R(\theta,\Delta_{n,i}^{(\frac{1}{2}-\beta)r\wedge 1},X_{t_i}) = R(\theta,\Delta_{n,i},X_{t_i}). \tag{71}$$

On the first probability here above we have used Tchebychev inequality and the fourth point of Lemma 1, for $|\Delta X_i^J| \geq \frac{\Delta_{n,i}^{\beta}}{2}$ the fact that the intensity of jumps is finite and therefore the probability to have at least one jump bigger than $\frac{\Delta_{n,i}^{\beta}}{2}$ can be computed and it is of order Δ_n. Moreover, by the arbitrariness of $r > 1$, we get that $R(\theta,\Delta_{n,i}^{(\frac{1}{2}-\beta)r},X_{t_i})$ is negligible compared to $R(\theta,\Delta_{n,i},X_{t_i})$.
From Holder inequality, the polynomial growth of f, the third point of Lemma 2 and (71) it follows that the right hand side of (70) goes to 0 in norm 1 and so in probability. The proposition is therefore proved. □

We now prove Proposition 6, that is a consequence of Lemma 3.

A.1.2 Proof of Proposition 6

Proof. In order to show that the first convergence holds, we define $s_i^n := \frac{1}{T_n}f(X_{t_i},\theta)\left(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i})\right)^2\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})1_{\left\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\right\}}$. From Lemma 9 in [14], if we show that

$$\sum_{i=0}^{n-1}\mathbb{E}_i[s_i^n] \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} f(x,\theta)a^2(x,\sigma_0)\pi(dx) \quad \text{and} \quad \sum_{i=0}^{n-1}\mathbb{E}_i[(s_i^n)^2] \to 0,$$

then the proposition is proved. We observe that (23) yields

$$\sum_{i=0}^{n-1}\mathbb{E}_i[s_i^n] = \frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)a^2(X_{t_i},\sigma_0)1_{\left\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\right\}} + \frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)R(\theta_0,\Delta_{n,i}^{\beta},X_{t_i}).$$

The first term here above converges in probability to $\int_{\mathbb{R}} f(x,\theta)a^2(x,\sigma_0)\pi(dx)$ as a consequence of the first point of Proposition 5, while the second is upper bounded by
$\Delta_n^\beta \frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)R(\theta_0,1,X_{t_i})$, which converges to zero in norm 1 (and so in probability) by the polynomial growth of both R and f and the third point of Lemma 2. Moreover, using (24) and the fact that $\frac{1}{T_n} = O(\frac{1}{n\Delta_n})$, we have

$$\sum_{i=0}^{n-1} |\mathbb{E}_i[(s_i^n)^2]| \leq \frac{1}{(n\Delta_n)^2}\sum_{i=0}^{n-1}(f(X_{t_i},\theta))^2 R(\theta_0,\Delta_{n,i}^2,X_{t_i}) \leq \frac{1}{n^2}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)R(\theta_0,1,X_{t_i}),$$

which goes to zero in norm 1 and so in probability for $n\to\infty$ as a consequence of the polynomial growth of both R and f and the third point of Lemma 2. The first point is therefore proved. In order to show the second point of Proposition 6 is enough to act on the sequence $\tilde{s}_i^n := \frac{1}{n}\frac{f(X_{t_i},\theta)}{\Delta_{n,i}}(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}$ exactly like we have just did here above, with the only difference that the third point of Proposition 5 has to be applied instead of the first one. □

A.1.3 Proof of Lemma 3

Proof. Replacing $m(\mu,\sigma,X_{t_i})$ with its development (31) and using the dynamic (3) of X we have

$$X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}) = \int_{t_i}^{t_{i+1}} b(X_s,\mu_0)ds + \int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s + \int_{t_i}^{t_{i+1}}\int_{\mathbb{R}} z\gamma(X_{s^-})\tilde{\mu}(ds,dz) + R(\theta,\Delta_{n,i},X_{t_i}) =$$

$$=: \int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s + B_{i,n}. \tag{72}$$

In order to prove (23) we start considering

$$\mathbb{E}_i[(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] = \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] +$$

$$+\mathbb{E}_i[(B_{i,n})^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] + 2\mathbb{E}_i[B_{i,n}(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})].$$

Now the first term on the right hand side here above is

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2(\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})-1)] =$$

$$= \Delta_{n,i}a^2(X_{t_i},\sigma_0) + \mathbb{E}_i[\int_{t_i}^{t_{i+1}}[a^2(X_s,\sigma_0) - a^2(X_{t_i},\sigma_0)]ds] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2(\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})-1)]. \tag{73}$$

Moreover,

$$|\mathbb{E}_i[\int_{t_i}^{t_{i+1}}[a^2(X_s,\sigma_0) - a^2(X_{t_i},\sigma_0)]ds] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2(\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})-1)]| \leq$$

$$\leq \int_{t_i}^{t_{i+1}} \mathbb{E}_i[|2a\partial_x a(X_u,\sigma_0)||X_s - X_{t_i}|]ds + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^{2p}]^{\frac{1}{p}}\mathbb{E}_i[1_{\{|\Delta X_i|\geq\Delta_{n,i}^\beta\}}]^{\frac{1}{q}}, \tag{74}$$

where $X_u \in (X_s, X_{t_i})$ and we have used Holder inequality and the definition of φ, that is equal to 1 for $|\Delta X_i| < \Delta_{n,i}^\beta$.
Using Cauchy-Schwartz inequality and the second point of Lemma 1 on the first term of (74) and Burkholder-Davis-Gundy inequality and (71) on the second we have that the right hand side is upper bounded by

$$\int_{t_i}^{t_{i+1}} R(\theta_0,\Delta_{n,i}^{\frac{1}{2}},X_{t_i})ds + R(\theta_0,\Delta_{n,i},X_{t_i})R(\theta_0,\Delta_{n,i}^{\frac{1}{q}},X_{t_i}) \leq R(\theta_0,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}) + R(\theta_0,\Delta_{n,i}^{2-\epsilon},X_{t_i}),$$

where we have taken q next to 1.
Replacing in (73) we get

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] = \Delta_{n,i}a^2(X_{t_i},\sigma_0) + R(\theta_0,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}). \tag{75}$$

Now we evaluate the contribution of $(B_{i,n})^2$:

$$\mathbb{E}_i[(B_{i,n})^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] \le c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} b(X_s,\mu_0)ds)^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})]+$$

$$+c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}}\gamma(X_{s^-})z\tilde{\mu}(ds,dz))^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] + R(\theta,\Delta_{n,i}^2,X_{t_i}) \le c\Delta_{n,i}\int_{t_i}^{t_{i+1}}\mathbb{E}_i[b^2(X_s,\mu)]ds+$$

$$+R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i}) = R(\theta_0,\Delta_{n,i}^2,X_{t_i}) + R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i}) = R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i}), \tag{76}$$

where we have used Jensen inequality, Lemma 5 and the fact that $R(\theta_0,\Delta_{n,i}^2,X_{t_i})$ is always negligible compared to $R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i})$ since $2>1+2\beta$.
We observe that (75) still holds with 1 instead of φ (see (73)). Using it, Cauchy-Schwartz inequality and (76) it follows

$$\mathbb{E}_i[B_{i,n}(\int_{t_i}^{t_{i+1}}a(X_s,\sigma_0)dW_s)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] \le c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_s,\sigma_0)dW_s)^2]^{\frac{1}{2}}\mathbb{E}_i[(B_{i,n})^2\varphi^2_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})]^{\frac{1}{2}} \le$$

$$\le R(\theta_0,\Delta_{n,i},X_{t_i})^{\frac{1}{2}}R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i})^{\frac{1}{2}} = R(\theta_0,\Delta_{n,i}^{1+\beta},X_{t_i}). \tag{77}$$

From (72), (75) - (77) it follows (23).

Concerning (24), we have

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_s,\sigma_0)dW_s)^4\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)] = \mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_s,\sigma_0)dW_s)^4]+\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_s,\sigma_0)dW_s)^4(\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)-1)]. \tag{78}$$

Using Holder inequality and the definition of φ we have that the second term here above is upper bounded by

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_s,\sigma_0)dW_s)^{4p}]^{\frac{1}{p}}\mathbb{P}_i(|\Delta X_i|\ge\Delta_{n,i}^\beta)^{\frac{1}{q}} \le R(\theta_0,\Delta_{n,i}^2,X_{t_i})R(\theta_0,\Delta_{n,i}^{\frac{1}{q}},X_{t_i}) = R(\theta_0,\Delta_{n,i}^{3-\epsilon},X_{t_i}), \tag{79}$$

where we have used BDG inequality, (71) and we have taken q next to 1. Moreover,

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_s,\sigma_0)dW_s)^4] = \mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_{t_i},\sigma_0)dW_s)^4]+\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^4]+ \tag{80}$$

$$+\sum_{j=1}^3\binom{4}{j}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_{t_i},\sigma_0)dW_s)^j(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^{4-j}].$$

Since the expected value of the fourth moment of the gaussian law is known we have

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_{t_i},\sigma_0)dW_s)^4] = 3\Delta_{n,i}^2a^4(X_{t_i},\sigma_0). \tag{81}$$

On the second term of the right hand side of (80) we use again BDG inequality to get

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^4] \le c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]^2ds)^2] \le$$

$$\le c\Delta_{n,i}\int_{t_i}^{t_{i+1}}\|\partial_x a\|_\infty^4\,\mathbb{E}_i[|X_s-X_{t_i}|^4]ds \le c\Delta_{n,i}\int_{t_i}^{t_{i+1}}|s-t_i|(1+|X_{t_i}|^4)ds \le R(\theta_0,\Delta_{n,i}^3,X_{t_i}), \tag{82}$$

where we have also used Jensen inequality and the second point of Lemma 1.
Concerning the last term in the right hand side of (80), from Holder inequality it is upper bounded by

$$\sum_{j=1}^3\binom{4}{j}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_{t_i},\sigma_0)dW_s)^{jp_1}]^{\frac{1}{p_1}}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^{(4-j)p_2}]^{\frac{1}{p_2}}.$$

Now we take $p_1=\frac{4}{j}$ and so $p_2=\frac{4}{4-j}$. Therefore, using also (81) and (82), the expression here above is

$$\sum_{j=1}^3\binom{4}{j}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_{t_i},\sigma_0)dW_s)^4]^{\frac{j}{4}}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^4]^{\frac{4-j}{4}} \le$$

$$\leq \sum_{j=1}^{3}\binom{4}{j}R(\theta_0,\Delta_{n,i}^2,X_{t_i})^{\frac{j}{4}}R(\theta_0,\Delta_{n,i}^3,X_{t_i})^{\frac{4-j}{4}} \leq \sum_{j=1}^{3}\binom{4}{j}R(\theta_0,\Delta_{n,i}^{3-\frac{j}{4}},X_{t_i}) = R(\theta_0,\Delta_{n,i}^{\frac{9}{4}},X_{t_i}), \quad (83)$$

since when $j=1$ and $j=2$ we get terms that are negligible if compared to $R(\theta_0,\Delta_{n,i}^{\frac{9}{4}},X_{t_i})$.
Replacing (79), (81) - (83) in (78) it follows

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^4\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)] = 3\Delta_{n,i}^2 a^4(X_{t_i},\sigma_0) + R(\theta_0,\Delta_{n,i}^{\frac{9}{4}},X_{t_i}). \quad (84)$$

We now study the contribution of $B_{i,n}$. First,

$$\mathbb{E}_i[(B_{i,n})^4|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|] \leq c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} b(X_s,\mu_0)ds)^4|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|]+$$

$$+c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}}\gamma(X_{s^-})z\tilde{\mu}(ds,dz))^4|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|] + R(\theta,\Delta_{n,i}^4,X_{t_i}) \leq c\Delta_{n,i}^3\int_{t_i}^{t_{i+1}}\mathbb{E}_i[b^4(X_s,\mu_0)]ds+$$

$$+R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i}) = R(\theta_0,\Delta_{n,i}^4,X_{t_i}) + R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i}) = R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i}), \quad (85)$$

where we have used Jensen inequality, Lemma 5 and the fact that $R(\theta_0,\Delta_{n,i}^4,X_{t_i})$ is always negligible compared to $R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i})$ since $4>1+4\beta$.
Using (72) we have that

$$\mathbb{E}_i[(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^4|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|] = \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^4|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|]+\mathbb{E}_i[(B_{i,n})^4|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|]+ \quad (86)$$

$$+\sum_{j=1}^{3}\binom{4}{j}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^j(B_{i,n})^{4-j}|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|].$$

On the last term here above we act like we did in (77), using holder inequality and taking $p_1=\frac{4}{j}$. It follows, using also (80), (81), (82), (83) and (85), that it is upper bounded by

$$\sum_{j=1}^{3}\binom{4}{j}R(\theta_0,\Delta_{n,i}^2,X_{t_i})^{\frac{j}{4}}R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i})^{\frac{4-j}{4}} = \sum_{j=1}^{3}\binom{4}{j}R(\theta_0,\Delta_{n,i}^{1+4\beta+\frac{j}{4}(1-4\beta)},X_{t_i}). \quad (87)$$

Since we have chosen $\beta>\frac{1}{4}$, the terms in which $j=1,2$ are negligible compared to the one in which $j=3$ and so we get $R(\theta_0,\Delta_{n,i}^{\frac{7}{4}+\beta},X_{t_i})$. From (84)- (87) it follows (24).

In order to show (25) we start considering $B_{i,n}$:

$$|\mathbb{E}_i[B_{i,n}\varphi_{\Delta_{n,i}^{\beta}}^k(\Delta X_i)]| \leq R(\theta_0,\Delta_{n,i},X_{t_i}) + c\mathbb{E}_i[|\int_{t_i}^{t_{i+1}} b(X_s,\mu)ds|] + c\mathbb{E}_i[|\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}} z\gamma(X_{s^-})\tilde{\mu}(ds,dz)|] \leq$$

$$\leq R(\theta_0,\Delta_{n,i},X_{t_i}) + c\mathbb{E}_i[\int_{t_i}^{t_{i+1}}|b(X_s,\mu)|ds] + c\mathbb{E}_i[\int_{t_i}^{t_{i+1}}(\int_{\mathbb{R}}|z|F(z)dz)|\gamma(X_{s^-})|ds] \leq R(\theta_0,\Delta_{n,i},X_{t_i}), \quad (88)$$

having used the definition of $B_{i,n}$ given in (72), the boundedness of φ^k, the polynomial growth of both b and γ and the third point of Lemma 1.
Moreover,

$$|\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)\varphi_{\Delta_{n,i}^{\beta}}^k(\Delta X_i)]| = |\mathbb{E}_i[\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s]+\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)(\varphi_{\Delta_{n,i}^{\beta}}^k(\Delta X_i)-1)]| \leq$$

$$\leq R(\theta_0,\Delta_{n,i}^{\frac{1}{2}},X_{t_i})\mathbb{E}_i[1_{\{|\Delta X_i|\geq\Delta_{n,i}^{\beta}\}}]^{\frac{1}{q}} \leq R(\theta_0,\Delta_{n,i}^{\frac{3}{2}-\epsilon},X_{t_i}), \quad (89)$$

where we have used (71) and taken q next to 1. From the inequality here above and (88) it follows (25).

Concerning (26); we have

$$\mathbb{E}_i[|X_{t_{i+1}}-m(\mu,\sigma,X_{t_i})|^k|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|^{k'}] \leq c\mathbb{E}_i[|\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s|^k|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|^{k'}]+c\mathbb{E}_i[|B_{i,n}|^k|\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)|^{k'}] \leq$$

$$\leq R(\theta_0,\Delta_{n,i}^{\frac{k}{2}},X_{t_i}) + R(\theta_0,\Delta_{n,i}^k,X_{t_i}) + R(\theta_0,\Delta_{n,i}^{1+k\beta},X_{t_i}) = R(\theta_0,\Delta_{n,i}^{\frac{k}{2}\wedge(1+k\beta)},X_{t_i}),$$

where we have used on the first term here above the fact that φ is bounded and BDG inequality while on the second we have acted like we did in (76) or (85), with q that this time is equal to k.

We now want to show the fifth and last point of the lemma. Using (72) we have

$$(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))^3 = \sum_{j=0}^{3} \binom{3}{j} (\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^j B_{i,n}^{3-j}.$$

Therefore

$$\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))^3 \varphi'_{\Delta_{n,i}^\beta}(\Delta X_i)] = \sum_{j=0}^{3} \binom{3}{j} \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^j B_{i,n}^{3-j} \varphi'_{\Delta_{n,i}^\beta}(\Delta X_i)].$$

We observe that, for $j = 3$, it is

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^3 \varphi'_{\Delta_{n,i}^\beta}(\Delta X_i)] = \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^3] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^3 (\varphi'_{\Delta_{n,i}^\beta}(\Delta X_i) - 1)] \leq$$

$$\leq c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_{t_i}, \sigma_0) dW_s)^3] + c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} [a(X_s, \sigma_0) - a(X_{t_i}, \sigma_0)] dW_s)^3] + R(\theta_0, \Delta_{n,i}^{\frac{5}{2}-\epsilon}, X_{t_i}).$$

We remark that the first term here above is centered while on the second we can act like we did on (82) (still with an exponent that is 3 instead of 4), obtaining

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} [a(X_s, \sigma_0) - a(X_{t_i}, \sigma_0)] dW_s)^3] \leq R(\theta_0, \Delta_{n,i}^{\frac{5}{2}}, X_{t_i}).$$

For $j = 0$, instead, we get a term on which we act like we did in (76) or (85), with q that this time is equal to 3 and so we can upper bound it with $R(\theta_0, \Delta_{n,i}^{(1+3\beta)\wedge 3}, X_{t_i})$.
For $j = 1$ and $j = 2$ we use Holder inequality, getting

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^j B_{i,n}^{3-j} \varphi'_{\Delta_{n,i}^\beta}(\Delta X_i)] \leq \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^{jp}]^{\frac{1}{p}} \mathbb{E}_i[B_{i,n}^{(3-j)q} \varphi'^{q}_{\Delta_{n,i}^\beta}(\Delta X_i)]^{\frac{1}{q}} \leq$$

$$\leq \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^3]^{\frac{j}{3}} \mathbb{E}[B_{i,n}^3 \varphi'^{\frac{3}{3-j}}_{\Delta_{n,i}^\beta}(\Delta X_i)]^{1-\frac{j}{3}} = R(\theta_0, \Delta_{n,i}^{\frac{3}{2}\frac{j}{3}}, X_{t_i}) R(\theta_0, \Delta_{n,i}^{(1+3\beta)(1-\frac{j}{3})}, X_{t_i}) =$$

$$= R(\theta_0, \Delta_{n,i}^{1+3\beta+j(\frac{1}{6}-\beta)}, X_{t_i}).$$

Now, since $\beta > \frac{1}{4} > \frac{1}{6}$, the term we get for $j = 1$ is negligible compared to the one we get for $j = 2$, which is $R(\theta_0, \Delta_{n,i}^{\frac{4}{3}+\beta}, X_{t_i})$. In conclusion we have

$$\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))^3 \varphi'_{\Delta_{n,i}^\beta}(\Delta X_i)] = R(\theta_0, \Delta_{n,i}^{\frac{5}{2}-\epsilon}, X_{t_i}) + R(\theta_0, \Delta_{n,i}^{(1+3\beta)\wedge 3}, X_{t_i}) +$$

$$+ R(\theta_0, \Delta_{n,i}^{\frac{4}{3}+\beta}, X_{t_i}) = R(\theta_0, \Delta_{n,i}^{\frac{4}{3}+\beta}, X_{t_i}),$$

since we can always find an $\epsilon > 0$ for which $\frac{5}{2} - \epsilon > 1 + 3\beta > \frac{4}{3} + \beta$. The result follows. □

A.1.4 Proof of Lemma 4

Proof. We first of all observe that, for all $k \geq 1$, $|\varphi'_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|^k$ is different from 0 only if $|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]$. Recalling that from its definition (72) $B_{i,n} = \int_{t_i}^{t_{i+1}} b(X_s, \mu) ds + \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}} z\gamma(X_{s^-}) \tilde{\mu}(ds, dz) + R(\theta, \Delta_{n,i}, X_{t_i})$, it follows

$$\mathbb{E}[|X_{t_{i+1}}^\theta - m(\mu, \sigma, X_{t_i})|^p |\varphi'_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|^k] \leq c\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s^\theta, \sigma) dW_s|^p 1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\}}] +$$

$$+ c\mathbb{E}[|B_{i,n}|^p 1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\}}]. \tag{90}$$

On the first term here above we use Holder inequality, (50) and (71), remarking that $\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\} \subset \left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \geq \Delta_{n,i}^\beta\right\}$. We get it is upper bounded by

$$\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s^\theta, \sigma) dW_s|^{pp_1}]^{\frac{1}{p_1}} \mathbb{E}[1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\}}]^{\frac{1}{p_2}} \leq R(\theta, \Delta_{n,i}^{\frac{p}{2}}, X_{t_i}) R(\theta, \Delta_{n,i}, X_{t_i})^{\frac{1}{p_2}} = R(\theta, \Delta_{n,i}^{\frac{p}{2}+1-\epsilon}, X_{t_i}),$$

for all $\epsilon > 0$. In the last inequality we have taken p_1 big and p_2 next to 1.
We now study the second term of (90). From the definition of $B_{i,n}$ given here above, Holder inequality, the polynomial growth of b and still (71) we get that the second term of (90) is upper bounded by $R(\theta, \Delta_{n,i}^{p+1-\epsilon}, X_{t_i}) + \mathbb{E}[|\Delta X_i^J|^p 1_{\left\{|X^\theta_{t_{i+1}} - X^\theta_{t_i}| \in [\Delta^\beta_{n,i}, 2\Delta^\beta_{n,i}]\right\}}]$.
We now recall that $\Delta X_i^c = (X^\theta_{t_{i+1}} - X^\theta_{t_i}) - \Delta X_i^J$ and so when $|X^\theta_{t_{i+1}} - X^\theta_{t_i}| \leq 2\Delta^\beta_{n,i}$ and $|\Delta X_i^J| \geq 4\Delta^\beta_{n,i}$, then $|\Delta X_i^c|$ must be more than $2\Delta^\beta_{n,i}$. Hence

$$\mathbb{E}[|\Delta X_i^J|^p 1_{\left\{|X^\theta_{t_{i+1}} - X^\theta_{t_i}| \in [\Delta^\beta_{n,i}, 2\Delta^\beta_{n,i}]\right\}}] \leq \mathbb{E}[|\Delta X_i^J|^p 1_{\left\{|X^\theta_{t_{i+1}} - X^\theta_{t_i}| \in [\Delta^\beta_{n,i}, 2\Delta^\beta_{n,i}], |\Delta X_i^J| \geq 4\Delta^\beta_{n,i}\right\}}] +$$

$$+\mathbb{E}[|\Delta X_i^J|^p 1_{\left\{|X^\theta_{t_{i+1}} - X^\theta_{t_i}| \in [\Delta^\beta_{n,i}, 2\Delta^\beta_{n,i}], |\Delta X_i^J| \leq 4\Delta^\beta_{n,i}\right\}}] \leq c\mathbb{E}[|\Delta X_i^J|^{pp_1}]^{\frac{1}{p_1}} \mathbb{P}(|\Delta X_i^c| \geq 2\Delta^\beta_{n,i})^{\frac{1}{p_2}} +$$

$$+c\Delta^{\beta p}_{n,i} \mathbb{P}(|X^\theta_{t_{i+1}} - X^\theta_{t_i}| \in [\Delta^\beta_{n,i}, 2\Delta^\beta_{n,i}], |\Delta X_i^J| \leq 4\Delta^\beta_{n,i}) \leq \tag{91}$$

$$\leq R(\theta, \Delta_{n,i}^{\frac{1}{p_1}}, X_{t_i}) R(\theta, \Delta_{n,i}^{\frac{(\frac{1}{2}-\beta)r}{p_2}}, X_{t_i}) + R(\theta, \Delta_{n,i}^{1+\beta p}, X_{t_i}) = R(\theta, \Delta_{n,i}^{(\frac{1}{2}-\beta)r-\epsilon}, X_{t_i}) + R(\theta, \Delta_{n,i}^{1+\beta p}, X_{t_i}) = R(\theta, \Delta_{n,i}^{1+\beta p}, X_{t_i}),$$

where we have used Kunita inequality (for $pp_1 \geq 2$, that holds since we take p_1 big and p_2 next to 1), Tchebyschev inequality as we did in (71) on the first term and still (71) on the second. Moreover we have used that, by the arbitrariness of $r > 0$, we can always find r and ϵ such that $(\frac{1}{2} - \beta)r - \epsilon > 1 + \beta p$. The result follows. □

A.1.5 Proof of Lemma 5

Proof. The case $q \geq 2$ has already been proved in Lemma 10 of [2] so, we are going to focus on the case $q \in [1, 2)$.
For all $n \in \mathbb{N}$ and $i \in \mathbb{N}$ we define the set on which all the jumps of L on the interval $(t_i, t_{i+1}]$ are small:

$$N_n^i := \left\{ |\Delta L_s| \leq \frac{4\Delta^\beta_{n,i}}{\gamma_{min}}; \quad \forall s \in (t_i, t_{i+1}] \right\}.$$

We split the jumps on $N_{i,n}$ and its complementary, getting

$$\mathbb{E}_i[|\Delta X_i^J \varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}} - X_{t_i})|^q 1_{N_n^i}] + \mathbb{E}_i[|\Delta X_i^J \varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}} - X_{t_i})|^q 1_{(N_n^i)^c}]. \tag{92}$$

We now observe that, by the definition of N_n^i, the first term here above is upper bounded by

$$\mathbb{E}_i[|\int_{t_i}^{t_{i+1}} \int_{|z| \leq \frac{4\Delta^\beta_{n,i}}{\gamma_{min}}} z\, \gamma(X_{s^-}) \tilde{\mu}(ds, dz)|^q + |\int_{t_i}^{t_{i+1}} \int_{|z| \geq \frac{4\Delta^\beta_{n,i}}{\gamma_{min}}} |z|\, |\gamma(X_{s^-})| \bar{\mu}(ds, dz)|^q].$$

From our assumptions on the jump density the second term here above is upper bounded by a $R(\theta, \Delta^q_{n,i}, X_{t_i})$ function while on the first one we use Lemma 2.1.5 of [21]. We can therefore upper bound it with

$$c\mathbb{E}_i[\int_{t_i}^{t_{i+1}} \int_{|z| \leq \frac{4\Delta^\beta_{n,i}}{\gamma_{min}}} |z|^q |\gamma(X_{s^-})|^q \bar{\mu}(ds, dz)] \leq R(\theta, \Delta_{n,i}^{1+\beta q}, X_{t_i}),$$

having used again that $\bar{\mu}(ds, dz) = F(dz)ds$ and Assumption 4 on F. It follows

$$\mathbb{E}_i[|\Delta X_i^J \varphi_{\Delta^\beta_{n,i}}(\Delta X_i)|^q 1_{N_n^i}] \leq R(\theta, \Delta^q_{n,i}, X_{t_i}) + R(\theta, \Delta_{n,i}^{1+\beta q}, X_{t_i}) = R(\theta, \Delta^q_{n,i}, X_{t_i}).$$

Regarding the second term of (92), we have that $|\Delta X_i^J| \leq |\Delta X_i| + |\Delta X_i^c|$ and, as we have already remarked several times, by the definition of φ it is different from zero only if $|\Delta X_i|^q \leq c\Delta^{\beta q}_{n,i}$. It follows

$$\mathbb{E}_i[|\Delta X_i|^q |\varphi_{\Delta^\beta_{n,i}}(\Delta X_i)|^q 1_{(N_n^i)^c}] \leq c\Delta^{\beta q}_{n,i} \mathbb{P}_i((N_n^i)^c) \leq R(\theta, \Delta_{n,i}^{1+\beta q}, X_{t_i}),$$

where the last inequality is a consequence of the following:

$$\mathbb{P}_i((N_n^i)^c) = \mathbb{P}_i(\exists s \in (t_i, t_{i+1}] : |\Delta L_s| > \frac{4\Delta^\beta_{n,i}}{\gamma_{min}}) \leq c \int_{t_i}^{t_{i+1}} \int_{\frac{4\Delta^\beta_{n,i}}{\gamma_{min}}}^{\infty} F(z) dz ds \leq c\Delta_{n,i}.$$

In the same way

$$\mathbb{E}_i[|\Delta X_i^c|^q |\varphi_{\Delta^\beta_{n,i}}(\Delta X_i)|^q 1_{(N_n^i)^c}] \leq c\Delta_{n,i}^{\frac{1}{2}q-\epsilon} \Delta_{n,i}^{1-\epsilon}(1 + |X_{t_i}|^c) = R(\theta, \Delta_{n,i}^{1+\frac{1}{2}q}, X_{t_i}).$$

Putting all pieces together we have

$$\mathbb{E}_i[|\Delta X_i^J|^q |\varphi_{\Delta^\beta_{n,i}}(\Delta X_i)|^q] \leq R(\theta, \Delta^q_{n,i}, X_{t_i}),$$

that is the result we wanted remarking that, for $q \in [1, 2)$, $1 + \beta q > q$ and so $\Delta^q_{n,i} = \Delta_{n,i}^{q \wedge (1+\beta q)}$. □

A.1.6 Proof of Proposition 7

Proof. We want to prove the tightness of the sequence $S_n(\theta)$. Since the sum of tight sequences is still tight, we show the tightness of the sequence $S_{n1}(\theta)$ and $S_{n2}(\theta)$ which are such that $S_n(\theta) = S_{n1}(\theta) + S_{n2}(\theta)$:

$$S_{n1}(\theta) := \frac{1}{T_n}\sum_{i=0}^{n-1}(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})g_{i,n}(\theta, X_{t_i}) \quad \text{and}$$

$$S_{n2}(\theta) := \frac{1}{T_n}\sum_{i=0}^{n-1}(m(\mu_0, \sigma_0, X_{t_i}) - m(\mu, \sigma, X_{t_i}))\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})g_{i,n}(\theta, X_{t_i}).$$

We prove that $S_{n1}(\theta)$ is tight using Kolmogorov criterion, we therefore want to show that inequalities analogous to (46) and (47) hold. Starting with the proof of (47) we have that, using Burkholder and Jensen inequality,

$$\mathbb{E}[|S_{n1}(\theta_1) - S_{n1}(\theta_2)|^m] \leq \frac{cn^{\frac{m}{2}-1}}{(n\Delta_n)^m}\sum_{i=0}^{n-1}\mathbb{E}[|(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))|^m|\varphi^m_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|g_{i,n}(\theta_1, X_{t_i}) - g_{i,n}(\theta_2, X_{t_i})|^m]. \tag{93}$$

We observe that, using finite-increments theorem and the assumption on the derivatives of $g_{i,n}$ with respect to the parameters, it is

$$|g_{i,n}(\theta_1, X_{t_i}) - g_{i,n}(\theta_2, X_{t_i})|^m \leq |R(\theta, 1, X_{t_i})|^m|\mu_1 - \mu_2|^m + |R(\theta, 1, X_{t_i})|^m|\sigma_1 - \sigma_2|^m \tag{94}$$

where actually the function R is computed in a point $\tilde{\theta} := (\tilde{\mu}, \tilde{\sigma})$, with $\tilde{\mu} \in (\mu_1, \mu_2)$ and $\tilde{\sigma} \in (\sigma_1, \sigma_2)$ but, since the property (8) of R is uniform in θ, we have chosen to write it simply as $R(\theta, 1, X_{t_i})$. Replacing (94) in (93) and using the fourth point of Lemma 3 it follows

$$\mathbb{E}[|S_{n1}(\theta_1) - S_{n1}(\theta_2)|^m] \leq \frac{cn^{\frac{m}{2}-1}}{(n\Delta_n)^m}n\Delta_n^{\frac{m}{2}\wedge(1+m\beta)}(|\mu_1 - \mu_2|^m + |\sigma_1 - \sigma_2|^m) \leq \frac{c}{(n\Delta_n)^{\frac{m}{2}}}\Delta_n^{0\wedge(1+m\beta-\frac{m}{2})}(|\mu_1 - \mu_2|^m + |\sigma_1 - \sigma_2|^m),$$

with $\frac{1}{(n\Delta_n)^{\frac{m}{2}}}\Delta_n^{0\wedge(1+m\beta-\frac{m}{2})} < c$ because $n\Delta_n$ is lower bounded by a constant and we can always find an $m \geq 2$ for which $1 + m\beta - \frac{m}{2} > 0$ since $\beta \in (\frac{1}{4}, \frac{1}{2})$. Hence, (47) is proved.
Acting exactly in the same way but using this time the control on $g_{i,n}$ instead of on its derivatives we have also an estimation for S_{n1} analogous to (46); the tightness of S_{n1} follows.
Concerning S_{n2} we observe that, for $\vartheta = \mu$ and $\vartheta = \sigma$, it is

$$|\partial_\vartheta S_{n2}(\theta)| \leq c|\partial_\vartheta m(\mu, \sigma, X_{t_i})||g_{i,n}(\theta, X_{t_i})| + c|m(\mu_0, \sigma_0, X_{t_i}) - m(\mu, \sigma, X_{t_i})||\partial_\vartheta g_{i,n}(\theta, X_{t_i})|.$$

From the controls we have assumed on $g_{i,n}$ and its derivatives, the finite-increments theorem and the first and the second point of Proposition 8 we have $|\partial_\vartheta S_{n2}(\theta)| \leq R(\theta, \Delta_{n,i}, X_{t_i})$. Therefore for both $\vartheta = \mu$ and $\vartheta = \sigma$, using also that $\frac{1}{T_n} = O(\frac{1}{n\Delta_n})$, we get

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta S_{n2}(\theta)|] \leq \sup_n \frac{c}{T_n}\sum_{i=0}^{n-1}\mathbb{E}[\sup_{\mu,\sigma}|R(\theta, \Delta_{n,i}, X_{t_i})] \leq c.$$

The tightness of S_{n2} (and therefore of S_n) follows.

□

A.2 Proof of derivatives of m and m_2

In order to prove the developments and the bounds on the derivatives of m and m_2, the following lemmas will be useful. We point out that X_t^θ is $X_t^{\theta,x}$ and so the process starts in 0: $X_0^{\theta,x} = x$.

Lemma 14. *Suppose that Assumptions from 1 to 4 and A7 hold. Then, $\forall p \geq 2$ $\exists c > 0$: $\forall h \leq \Delta_n$ $\forall x$ we have*

$$\mathbb{E}[|\frac{\partial_\mu X_h^{\theta,x}}{h}|^p] \leq c(1 + |x|^c), \qquad \mathbb{E}[|\frac{\partial^2_\mu X_h^{\theta,x}}{h}|^p] \leq c(1 + |x|^c), \tag{95}$$

$$\mathbb{E}[|\frac{\partial_\sigma X_h^{\theta,x}}{h^{\frac{1}{2}}}|^p] \leq c(1 + |x|^c), \qquad \mathbb{E}[|\frac{\partial^2_{\sigma\mu} X_h^{\theta,x}}{h^{\frac{3}{2}}}|^p] \leq c(1 + |x|^c), \tag{96}$$

$$\mathbb{E}[|\frac{\partial^2_\sigma X_h^{\theta,x}}{h^{\frac{1}{2}}}|^p] \leq c(1 + |x|^c), \tag{97}$$

$$\mathbb{E}[|\frac{\partial^3_\sigma X^{\theta,x}_h}{h^{\frac{1}{2}}}|^p] \leq c(1+|x|^c), \qquad \mathbb{E}[|\frac{\partial^3_\mu X^{\theta,x}_h}{h}|^p] \leq c(1+|x|^c) \tag{98}$$

$$\mathbb{E}[|\frac{\partial^3_{\sigma\mu\sigma} X^{\theta,x}_h}{h^{\frac{3}{2}}}|^p] \leq c(1+|x|^c), \qquad \mathbb{E}[|\frac{\partial^3_{\mu\mu\sigma} X^{\theta,x}_h}{h^{\frac{3}{2}}}|^p] \leq c(1+|x|^c), \tag{99}$$

Proof. Lemma 14.
Inequalities (95) have already been proved in Lemma 9 of [2]. To show the first inequality of (96), we observe that the dynamic of the process $\partial_\sigma X^{\theta,x}$ is known (cf. [7], section 5):

$$\partial_\sigma X^{\theta,x}_h = \int_0^h \partial_x b(\mu, X^{\theta,x}_s)\partial_\sigma X^{\theta,x}_s ds + \int_0^h (\partial_x a(\sigma, X^{\theta,x}_s)\partial_\sigma X^{\theta,x}_s + \partial_\sigma a(\sigma, X^{\theta,x}_s))dW_s + \int_0^h \int_{\mathbb{R}} \partial_x \gamma(X^{\theta,x}_{s^-})\partial_\sigma X^{\theta,x}_s z\tilde{\mu}(dz,ds). \tag{100}$$

From now on, we will drop the dependence of the starting point in order to make the notation easier. Taking the L^p norm of (100) we get it is upper bounded by the sum of three terms. On the first one we use Jensen inequality and the fact that the derivatives of b with rapport to x are supposed bounded to obtain

$$\mathbb{E}[|\int_0^h (\partial_x b(\mu, X^\theta_s)\partial_\sigma X^\theta_s ds|^p] \leq h^{p-1}\int_0^h \mathbb{E}[|\partial_x b(\mu, X^\theta_s)|^p |\partial_\sigma X^\theta_s|^p]ds \leq ch^{p-1}\int_0^h \mathbb{E}[|\partial_\sigma X^\theta_s|^p]ds. \tag{101}$$

Let us now consider the stochastic integral. Using Burkholder-Davis-Gundy inequality we have

$$\begin{aligned}
&\mathbb{E}[|\int_0^h (\partial_x a(\sigma, X^\theta_s)\partial_\sigma X^\theta_s + \partial_\sigma a(\sigma, X^\theta_s))dW_s|^p] \leq c\mathbb{E}[|\int_0^h (\partial_x a(\sigma, X^\theta_s))^2(\partial_\sigma X^\theta_s)^2 ds|^{\frac{p}{2}}] + \\
&+c\mathbb{E}[|\int_0^h (\partial_\sigma a(\sigma, X^\theta_s))^2 ds|^{\frac{p}{2}}] \leq ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[|\partial_\sigma X^\theta_s|^p]ds + ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[(1+|X^\theta_s|^c)]ds \leq \\
&\leq ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[|\partial_\sigma X^\theta_s|^p]ds + ch^{\frac{p}{2}}(1+|x|^c),
\end{aligned} \tag{102}$$

where we have used Jensen inequality, the fact that, by A7, the derivatives of a with rapport to x are supposed bounded and those with rapport to σ have polynomial growth and the second point of Lemma 1.
We now consider the third term on the right hand side of (100), it can be estimated using Kunita inequality (cf. the Appendix of [21]):

$$\begin{aligned}
&\mathbb{E}[|\int_0^h \int_{\mathbb{R}} \partial_x \gamma(X^\theta_{s^-})\partial_\sigma X^\theta_s z\tilde{\mu}(dz,ds)|^p] \leq c\mathbb{E}[\int_0^h \int_{\mathbb{R}} |\partial_x \gamma(X^\theta_{s^-})\partial_\sigma X^\theta_s|^p |z|^p \bar{\mu}(dz,ds)] + \\
&+c\mathbb{E}[|\int_0^h \int_{\mathbb{R}} (\partial_x \gamma(X^\theta_{s^-})\partial_\sigma X^\theta_s)^2 z^2 \bar{\mu}(dz,ds)|^{\frac{p}{2}}] \leq c\int_0^h \mathbb{E}[|\partial_x \gamma(X^\theta_{s^-})|^p |\partial_\sigma X^\theta_s|^p](\int_{\mathbb{R}} |z|^p F(z)dz)ds + \\
&+c\mathbb{E}[|\int_0^h (\partial_x \gamma(X^\theta_{s^-})\partial_\sigma X^\theta_s)^2 (\int_{\mathbb{R}} z^2 F(z)dz)ds|^{\frac{p}{2}}] \leq c\int_0^h \mathbb{E}[|\partial_\sigma X^\theta_s|^p]ds + c\mathbb{E}[|\int_0^h (\partial_\sigma X^\theta_s)^2 ds|^{\frac{p}{2}}],
\end{aligned}$$

where in the last two inequalities we have just used the definition of the compensated measure $\bar{\mu}$,the third point of Assumption 4 and the fact that the derivatives of γ are supposed bounded. By the Jensen inequality we get

$$\mathbb{E}[|\int_0^h \int_{\mathbb{R}} \partial_x \gamma(X^\theta_{s^-})\partial_\sigma X^\theta_s z\tilde{\mu}(dz,ds)|^p] \leq c(1+h^{\frac{p}{2}-1})\int_0^h \mathbb{E}[|\partial_\sigma X^\theta_s|^p]ds. \tag{103}$$

From (101), (102) and (103), it follows

$$\mathbb{E}[|\partial_\sigma X^\theta_h|^p] \leq c(h^{p-1} + h^{\frac{p}{2}-1} + 1)\int_0^h \mathbb{E}[|\partial_\sigma X^\theta_s|^p]ds + ch^{\frac{p}{2}}(1+|x|^c).$$

Gronwall Lemma gives us

$$\mathbb{E}[|\partial_\sigma X^\theta_h|^p] \leq ch^{\frac{p}{2}}(1+|x|^c)e^{c(h^{p-1}+h^{\frac{p}{2}-1}+1)},$$

we therefore obtain the first inequality in (96). Concerning the second, we observe we can deduce the dynamic of the process $\partial_{\mu\sigma} X^\theta$ from (100). It is

$$\partial^2_{\mu\sigma} X^\theta_h = \int_0^h (\partial^2_x b(\mu, X^\theta_s)\partial_\sigma X^\theta_s \partial_\mu X^\theta_s + \partial^2_{\mu x} b(\mu, X^\theta_s)\partial_\sigma X^\theta_s + \partial_x b(\mu, X^\theta_s)\partial^2_{\sigma\mu} X^\theta_s)ds + \int_0^h (\partial^2_x a(\sigma, X^\theta_s)\partial_\sigma X^\theta_s \partial_\mu X^\theta_s +$$

$$+\partial^2_{\sigma x}a(\sigma,X^\theta_s)\partial_\mu X^\theta_s+\partial_x a(\sigma,X^\theta_s)\partial^2_{\sigma\mu}X^\theta_s)dW_s+\int_0^h\int_{\mathbb{R}}(\partial^2_x\gamma(X^\theta_{s^-})\partial_\sigma X^\theta_s\partial_\mu X^\theta_s+\partial_x\gamma(X^\theta_{s^-})\partial^2_{\sigma\mu}X^\theta_s)z\tilde{\mu}(ds,dz)$$

On the p-norm of the first integral we use Jensen inequality, the fact that the derivatives with respect to x are bounded and the estimation we have already proved on the L^p norm of the derivatives of our process with respect to μ and σ. We get it is upper bounded by

$$ch^{p-1}\int_0^h(\mathbb{E}[|\partial_\sigma X^\theta_s\partial_\mu X^\theta_s|^p]+\mathbb{E}[|\partial_\sigma X^\theta_s|^p]+\mathbb{E}[|\partial^2_{\sigma\mu}X^\theta_s|^p])ds\leq c(h^{\frac{5}{2}p}+h^{\frac{3}{2}p})(1+|x|^c)+ch^{p-1}\int_0^h\mathbb{E}[|\partial^2_{\sigma\mu}X^\theta_s|^p]ds,$$

having also used Holder inequality. Acting in the same way on the stochastic integral we get it is upper bounded by

$$ch^{\frac{p}{2}}(h^{\frac{3}{2}p}+h^p)(1+|x|^c)+ch^{\frac{p}{2}-1}\int_0^h\mathbb{E}[|\partial^2_{\sigma\mu}X^\theta_s|^p]ds,$$

while we upper bound the third term in the dynamic of $\partial^2_{\sigma\mu}X^\theta$, acting as we did in order to show (103), with

$$ch^{\frac{3}{2}p+1}(1+|x|^c)+c\int_0^h\mathbb{E}[|\partial^2_{\sigma\mu}X^\theta_s|^p]ds+ch^{2p}(1+|x|^c)+ch^{\frac{p}{2}-1}\int_0^h\mathbb{E}[|\partial^2_{\sigma\mu}X^\theta_s|^p]ds.$$

In total we have, not considering the negligible terms,

$$\mathbb{E}[|\partial^2_{\sigma\mu}X^\theta_h|^p]\leq ch^{\frac{3}{2}p}(1+|x|^c)+c(h^{p-1}+h^{\frac{p}{2}-1}+1)\int_0^h\mathbb{E}[|\partial^2_{\sigma\mu}X^\theta_s|^p]ds.$$

From Gronwall Lemma it follows the second inequality of (96), as we wanted.
We are left to show (97). Again, the dynamic of $\partial^2_\sigma X^\theta$ is known:

$$\partial^2_\sigma X^\theta_h=\int_0^h(\partial^2_x b(\mu,X^\theta_s)(\partial_\sigma X^\theta_s)^2+\partial_x b(\mu,X^\theta_s)\partial^2_\sigma X^\theta_s)ds+\int_0^h(\partial^2_x a(\sigma,X^\theta_s)(\partial_\sigma X^\theta_s)^2+2\partial^2_{\sigma x}a(\sigma,X^\theta_s)\partial_\sigma X^\theta_s+$$

$$+\partial_x a(\sigma,X^\theta_s)\partial^2_\sigma X^\theta_s+\partial^2_\sigma a(\sigma,X^\theta_s))dW_s+\int_0^h\int_{\mathbb{R}}(\partial^2_x\gamma(X^\theta_{s^-})(\partial_\sigma X^\theta_s)^2+\partial_x\gamma(X^\theta_{s^-})\partial^2_\sigma X^\theta_s)z\tilde{\mu}(ds,dz). \quad (104)$$

Acting exactly like we did for the estimation of the p-moments of the processes $\partial_\sigma X^\theta$ and $\partial^2_{\sigma\mu}X^\theta$ we get

$$\mathbb{E}[|\partial^2_\sigma X^\theta_h|^p]\leq c(1+|x|^c)(h^{2p}+h^{\frac{3}{2}p}+h^p+h^{\frac{p}{2}}+h^{p+1})+c(h^{p-1}+h^{\frac{p}{2}-1}+1)\int_0^h\mathbb{E}[|\partial^2_\sigma X^\theta_s|^p]ds. \quad (105)$$

Using Gronwall Lemma and remarking that the other terms are negligible compared to $ch^{\frac{p}{2}}(1+|x|^c)$, we obtain the result wanted.
Concerning the third derivatives, it is easy to see that, writing the dynamics of $\partial^3_\sigma X^\theta_h$, $\partial^3_\mu X^\theta_h$, $\partial^3_{\sigma\mu\sigma}X^\theta_h$ and $\partial^3_{\mu\mu\sigma}X^\theta_h$ the principal terms are such that their order are, respectively, $h^{\frac{1}{2}}$, h and twice $h^{\frac{3}{2}}$. Acting exactly like before, (98) and (99) follow. □

We are left to show one last proposition, before showing Propositions 8, 9 and 10:

Proposition 11. *Suppose that Assumptions 1 to 4 hold. Moreover suppose that $(Z_h)_h$ is a family of random variables such that $\mathbb{E}[|Z_h|^p|X^\theta_0=x]\leq c(1+|x|^c)$. Then $\forall k\geq 1\ \forall\epsilon>0$, we have*

$$\sup_{h\in[0,\Delta_n]}\mathbb{E}[|Z_h||\varphi^{(k)}_{h^\beta}(X^\theta_h-x)||X^\theta_0=x]=R(\theta,h^{1-\epsilon},x).$$

We have used $\varphi^{(k)}_{h^\beta}(y)$ in order to denote the k-th derivative $\varphi^{(k)}(\frac{y}{h^\beta})$.

Proof. Proposition 11.
Once again, $|\varphi^{(k)}_{h^\beta}(X^\theta_h-x)|$ is different from 0 only if $|X^\theta_h-x|\in[h^\beta,2h^\beta]$. We can therefore use Holder inequality (with p big and q next to 1) and (71) to get, $\forall h\in[0,\Delta_n]$,

$$\mathbb{E}[|Z_h||\varphi^{(k)}_{h^\beta}(X^\theta_h-x)||X^\theta_0=x]\leq\mathbb{E}[|Z_h|^p|X^\theta_0=x]^{\frac{1}{p}}\mathbb{E}[1_{\{|X^\theta_h-x|\in[h^\beta,2h^\beta]\}}|X^\theta_0=x]^{\frac{1}{q}}\leq R(\theta,h,x)^{\frac{1}{q}}=R(\theta,h^{1-\epsilon},x).$$

□

A.2.1 Proof of Proposition 8

Proof. The first point has already been showed in Proposition 8 of [2], we start proving the second. Since by the homogeneity of the equation m and m_2 depend only on the difference $t_{i+1} - t_i$ we can consider WLOG $\forall h \leq \Delta_n$

$$m(\mu, \sigma, x) := \frac{\mathbb{E}[X_h^\theta \varphi_{h^\beta}(X_h^\theta - X_0^\theta)|X_0^\theta = x]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - X_0^\theta)|X_0^\theta = x]} = \frac{\mathbb{E}[X_h^\theta \varphi_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]}. \tag{106}$$

Hence,

$$\partial_\sigma m(\mu, \sigma, x) = \frac{\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi_{h^\beta}(X_h^\theta - x)] + \mathbb{E}[X_h^\theta h^{-\beta}(\partial_\sigma X_h^\theta)\varphi'_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} - m(\mu, \sigma, x)\frac{\mathbb{E}[h^{-\beta}(\partial_\sigma X_h^\theta)\varphi'_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]}.$$

On the numerator of the second and third term we use Proposition 11 taking as Z_h, respectively, $X_h^\theta \frac{\partial_\sigma X_h^\theta}{h^{\frac{1}{2}}}$ and $\frac{\partial_\sigma X_h^\theta}{h^{\frac{1}{2}}}$. We get, remarking moreover that from (31) $m(\mu, \sigma, x)$ is $R(\theta, 1, x)$ and from Theorem 1 in [2] also the denominator is lower bounded for $|x| < |h|^{-k}$,

$$|\partial_\sigma m(\mu, \sigma, x)| \leq R(\theta, 1, x)|\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi_{h^\beta}(X_h^\theta - x)]| + R(\theta, h^{\frac{3}{2}-\beta-\epsilon}, x). \tag{107}$$

To estimate $|\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi_{h^\beta}(X_h^\theta - x)]|$ we replace the dynamic (100) of $\partial_\sigma X_h^\theta$. On the first integral we use Holder inequality and (101) to get

$$|\mathbb{E}[\int_0^h (\partial_x b(X_s^{\theta,x}, \mu)\partial_\sigma X_s^{\theta,x} ds \varphi_{h^\beta}(X_h^\theta - x)]| \leq (ch^{p-1}\int_0^h \mathbb{E}[|\partial_\sigma X_s^{\theta,x}|^p])^{\frac{1}{p}} \leq ch^{\frac{3}{2}}(1 + |x|^c) = R(\theta, h^{\frac{3}{2}}, x),$$

where in the last inequality we have also used the first inequality of (96). On $\int_0^h \partial_x a(X_s^{\theta,x}, \sigma)\partial_\sigma X_s^{\theta,x} dW_s$ we use again Holder inequality, (102) (considering only the estimation on its first term) and the first inequality of (96) to obtain

$$|\mathbb{E}[\int_0^h \partial_x a(X_s^{\theta,x}, \sigma)\partial_\sigma X_s^{\theta,x} dW_s \varphi_{h^\beta}(X_h^\theta - x)]| \leq R(\theta, h, x).$$

Concerning $|\mathbb{E}[\int_0^h \partial_\sigma a(X_s^{\theta,x}, \sigma) dW_s \varphi_{h^\beta}(X_h^\theta - x)]|$, we act on it like we did in (89), with $\partial_\sigma a$ instead of a. We therefore get $|\mathbb{E}[\int_0^h \partial_\sigma a(X_s^{\theta,x}, \sigma) dW_s \varphi_{h^\beta}(X_h^\theta - x)]| \leq R(\theta, h, x)$. To conclude the proof of this point we use on the jump part Holder inequality, (103) and the first inequality of (96). We get

$$|\mathbb{E}[\int_0^h \int_{\mathbb{R}} \partial_x \gamma(X_{s^-}^\theta)\partial_\sigma X_s^\theta z\tilde{\mu}(dz, ds)\varphi_{h^\beta}(X_h^\theta - x)]| \leq (c(1 + h^{\frac{p}{2}-1})\int_0^h \mathbb{E}[|\partial_\sigma X_s^\theta|^p] ds)^{\frac{1}{p}} \leq R(\theta, h^{(\frac{1}{p}+\frac{1}{2})\wedge 1}, x).$$

We can take $p = 2$, finding $|\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi_{h^\beta}(X_h^\theta - x)]| = R(\theta, h, x)$. Replacing it in (107) and observing that $\frac{3}{2} - \beta$ is always more than 1, it follows the second point of Proposition 8.
In order to prove the third and the fourth point we first of all need to compute the derivative of m_2 with respect to both the parameters. We can just write

$$\partial_\vartheta m_2(\mu, \sigma, x) = 2\frac{\mathbb{E}[(X_h^\theta - m(\mu, \sigma, x))(\partial_\vartheta X_h^\theta - \partial_\vartheta m(\mu, \sigma, x))\varphi_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} +$$

$$+\frac{\mathbb{E}[(X_h^\theta - m(\mu, \sigma, x))^2 h^{-\beta}(\partial_\vartheta X_h^\theta)\varphi'_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} - m_2(\mu, \sigma, x)h^{-\beta}\frac{\mathbb{E}[(\partial_\vartheta X_h^\theta)\varphi'_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} =: I_{1,\theta} + I_{2,\theta} + I_{3,\theta}.$$

We are going to show that, considering the derivatives with respect to both μ and σ, $I_{2,\theta}$ and $I_{3,\theta}$ are negligible compared to $I_{1,\theta}$. In order to prove it we use Theorem 1 of [2] on the denominator of $I_{2,\theta}$ and $I_{3,\theta}$, while on the numerator of $I_{2,\theta}$ we use Holder inequality, (95) if we consider the derivative with respect to μ or the first equation of (96) if we consider the derivative with respect to σ and Lemma 4. On the numerator of $I_{3,\theta}$ we use Proposition 11 and we remind that, as a consequence of Ad, m_2 is a $R(\theta, h, x)$ function. It follows

$$|I_{2,\mu} + I_{3,\mu}| \leq R(\theta, h^{2+\beta-\epsilon}, x) + R(\theta, h^{3-\beta-\epsilon}, x) = R(\theta, h^{2+\beta-\epsilon}, x); \tag{108}$$

$$|I_{2,\sigma} + I_{3,\sigma}| \leq R(\theta, h^{\frac{3}{2}+\beta-\epsilon}, x) + R(\theta, h^{\frac{5}{2}-\beta-\epsilon}, x) = R(\theta, h^{\frac{3}{2}+\beta-\epsilon}, x). \tag{109}$$

Concerning $I_{1,\theta}$, its numerator is

$$2\mathbb{E}[(X_h^\theta - m(\mu, \sigma, x))\partial_\theta X_h^\theta \varphi_{h^\beta}(X_h^\theta - x)] - 2\partial_\theta m(\mu, \sigma, x))\mathbb{E}[(X_h^\theta - m(\mu, \sigma, x))\varphi_{h^\beta}(X_h^\theta - x)] =: I_{1,1}^\theta + I_{1,2}^\theta.$$

From the first two points we have already proved of Proposition 8 and the third point of Lemma 3 we get

$$|I^{\mu}_{1,2}| \leq R(\theta, h^2, x), \qquad |I^{\sigma}_{1,2}| \leq R(\theta, h^2, x). \tag{110}$$

Now we consider $I^{\theta}_{1,1}$ and we act differently depending on if we are dealing with the derivative with rapport to μ or those with rapport to σ. We start studying $I^{\sigma}_{1,1}$. Using a notation analogous to the one used in the proof of Lemma 3, we set

$$X^{\theta}_h - m(\mu, \sigma, x) =: \int_0^h a(X^{\theta}_s, \sigma) dW_s + B_h$$

and so it turns we have

$$I^{\sigma}_{1,1} = 2\mathbb{E}[(\int_0^h a(X^{\theta}_s, \sigma) dW_s) \partial_{\sigma} X^{\theta}_h \varphi_{h^{\beta}}(X^{\theta}_h - x)] + 2\mathbb{E}[(B_h) \partial_{\sigma} X^{\theta}_h \varphi_{h^{\beta}}(X^{\theta}_h - x)]. \tag{111}$$

On the second term here above we use Cauchy-Schwartz inequality, control analogous to (76) and the first estimation in (96), getting

$$\mathbb{E}[(B_h)\partial_{\sigma} X^{\theta}_h \varphi_{h^{\beta}}(X^{\theta}_h - x)] \leq c\mathbb{E}[(B_h)^2 \varphi^2_{h^{\beta}}(X^{\theta}_h - x)]^{\frac{1}{2}} \mathbb{E}[(\partial_{\sigma} X^{\theta}_h)^2]^{\frac{1}{2}} \leq R(\theta, h^{1+2\beta}, x)^{\frac{1}{2}} R(\theta, h^{\frac{1}{2}}, x) = R(\theta, h^{1+\beta}, x). \tag{112}$$

To evaluate the first term on the right hand side of (111) we replace the dynamic (100) of $\partial_{\sigma} X^{\theta}_h$, isolating the principal term: $\partial_{\sigma} X^{\theta}_h := \int_0^h \partial_{\sigma} a(X^{\theta}_s, \sigma) dW_s + G_{\sigma}$. We get

$$\mathbb{E}[(\int_0^h a(X^{\theta}_s, \sigma) dW_s)(\int_0^h \partial_{\sigma} a(X^{\theta}_s, \sigma) dW_s)] + \mathbb{E}[(\int_0^h a(X^{\theta}_s, \sigma) dW_s)(\int_0^h \partial_{\sigma} a(X^{\theta}_s, \sigma) dW_s)(\varphi_{h^{\beta}}(X^{\theta}_h - x) - 1)] + \tag{113}$$

$$+ \mathbb{E}[(\int_0^h a(X^{\theta}_s, \sigma) dW_s) G_{\sigma} \varphi_{h^{\beta}}(X^{\theta}_h - x)].$$

On the first term here above we add and substract both $a(x, \sigma)$ and $\partial_{\sigma} a(x, \sigma)$ getting a main term and three terms of increments. We observe it holds the following estimation, using Cauchy-Schwartz inequality, (50) and the first point of Lemma 1

$$\mathbb{E}[(\int_0^h [a(X^{\theta}_s, \sigma) - a(x, \sigma)] dW_s)(\int_0^h \partial_{\sigma} a(X^{\theta}_s, \sigma) dW_s)] \leq \mathbb{E}[(\int_0^h [a(X^{\theta}_s, \sigma) - a(x, \sigma)] dW_s)^2]^{\frac{1}{2}} \mathbb{E}[(\int_0^h \partial_{\sigma} a(X^{\theta}_s, \sigma) dW_s)^2]^{\frac{1}{2}} \leq \tag{114}$$

$$\leq c\mathbb{E}[\int_0^h (X_s - x)^2 ds]^{\frac{1}{2}} R(\theta, h^{\frac{1}{2}}, x) \leq R(\theta, h, x) R(\theta, h^{\frac{1}{2}}, x) = R(\theta, h^{\frac{3}{2}}, x).$$

We can act in the same way considering the increments of $\partial_{\sigma} a$ or the term on which we have the increments of both a and $\partial_{\sigma} a$. It follows that the first term of (113) is

$$\mathbb{E}[(\int_0^h a(x, \sigma) dW_s)(\int_0^h \partial_{\sigma} a(x, \sigma) dW_s)] + R(\theta, h^{\frac{3}{2}}, x) = h\, a(x, \sigma) \partial_{\sigma} a(x, \sigma) + R(\theta, h^{\frac{3}{2}}, x). \tag{115}$$

On the second term of (113) we use Holder inequality twice (with p big and q next to 1), (50) twice and (71). We get it is upper bounded by

$$\mathbb{E}[|(\int_0^h a(X^{\theta}_s, \sigma) dW_s)(\int_0^h \partial_{\sigma} a(X^{\theta}_s, \sigma) dW s)|^p]^{\frac{1}{p}} \mathbb{E}[(\varphi_{h^{\beta}}(X^{\theta}_h - x) - 1)^q]^{\frac{1}{q}} \leq R(\theta, h, x) R(\theta, h, x)^{\frac{1}{q}} = R(\theta, h^{2-\epsilon}, x)$$

Concerning the third term of (113), we first of all use Cauchy-Schwartz inequality, the fact that φ is bounded in absolute value and (50) in order to estimate the stochastic integral while, to estimate the 2-norm of G_{σ}, we use (101), the estimation (102) about the negligible part of the stochastic integral and (103). It follows it is upper bounded by

$$R(\theta, h^{\frac{1}{2}}, x) c(h^{\frac{3}{2}} + h) = R(\theta, h^{\frac{3}{2}}, x), \tag{116}$$

where we have also used the first estimation (96) to estimate the expected value. From (112), (115) - (116) it follows

$$I^{\sigma}_{1,1} = h\, a(x, \sigma) \partial_{\sigma} a(x, \sigma) + R(\theta, h^{1+\beta}, x).$$

Using also (109) and (110) we get the development of $\partial_{\sigma} m_2$ we wanted.
We are left to prove the third point of Proposition 8. It means, comparing it with (108) and (110), to prove that $|I^{\mu}_{1,1}| \leq R(\theta, h^2, x)$. We observe that (111) still holds with the derivative with respect to μ

instead of those with respect to σ. We now recall that $B_h = \int_0^h b(X_s,\mu)ds + R(\theta,h,x) + \Delta X_h^J$ and we replace it in $\mathbb{E}[B_h(\partial_\mu X_s^\theta)\varphi_{h^\beta}(X_h^\theta - x)]$, getting

$$|\mathbb{E}[(\int_0^h b(X_s,\mu)ds)(\partial_\mu X_s^\theta)\varphi_{h^\beta}(X_h^\theta - x)] + R(\theta,h,x)\mathbb{E}[(\partial_\mu X_s^\theta)\varphi_{h^\beta}(X_h^\theta - x)] + \mathbb{E}[(\Delta X_h^J)(\partial_\mu X_s^\theta)\varphi_{h^\beta}(X_h^\theta - x)]| \leq \tag{117}$$
$$\leq R(\theta,h^2,x) + R(\theta,h^{1+\beta q},x)^{\frac{1}{q}} R(\theta,h,x) = R(\theta,h^2,x) + R(\theta,h^{2+\beta-\epsilon},x) = R(\theta,h^2,x),$$

where we have used Holder inequality (having taken p big and q next to 1), the fact that φ is bounded, the polynomial growth of b, the first estimation in (95) and Lemma 5 (in a non-conditional form). Concerning the first term of (111), we still replace the dynamic of $\partial_\mu X_s^\theta := \int_0^h \partial_\mu b(X_s^\theta,\mu)ds + G_\mu$, where G_μ is the negligible part in the dynamic of $\partial_\mu X_s^\theta$ and it is such that

$$\mathbb{E}[|G_\mu|^p] \leq c(1 + h^{\frac{p}{2}-1} + h^{p-1})\int_0^h \mathbb{E}[|\partial_\mu X_s^\theta|^p]ds \leq c(h^{p+1} + h^{\frac{3}{2}p} + h^{2p})(1+|x|^c)$$

(see Lemma 9 in [2]). We have also used the first inequality of (95). It follows $\mathbb{E}[|G_\mu|^p]^{\frac{1}{p}} \leq R(\theta,h^{1+\frac{1}{p}},x)$, for $p \geq 2$. The first term of (111) is therefore

$$\mathbb{E}[(\int_0^h a(X_s^\theta,\sigma)dW_s)(\int_0^h \partial_\mu b(X_s^\theta,\mu)ds)] + \mathbb{E}[(\int_0^h a(X_s^\theta,\sigma)dW_s)(\int_0^h \partial_\mu b(X_s^\theta,\mu)ds)(1-\varphi_{h^\beta}(X_h^\theta - x))] + \tag{118}$$
$$+\mathbb{E}[(\int_0^h a(X_s^\theta,\sigma)dW_s)G_\mu\varphi_{h^\beta}(X_h^\theta - x)].$$

Now on the first term here above we act like we did for the estimation of the derivative with respect to σ: we get

$$h\,\partial_\mu b(x,\mu)\mathbb{E}[\int_0^h a(X_s^\theta,\sigma)dW_s] + \mathbb{E}[(\int_0^h a(X_s^\theta,\sigma)dW_s)(\int_0^h [b(X_s^\theta,\mu) - b(x,\mu)]ds)]$$

Now we observe that the first expected value is 0, while on the others we can use Cauchy-Schwartz inequality, (50) and we estimate the increments like in (114):

$$\mathbb{E}[(\int_0^h a(X_s^\theta,\sigma)dW_s)(\int_0^h [b(X_s^\theta,\mu) - b(x,\mu)]ds)] \leq \mathbb{E}[(\int_0^h a(X_s^\theta,\sigma)dW_s)^2]^{\frac{1}{2}}\mathbb{E}[(\int_0^h [b(X_s^\theta,\mu) - b(x,\mu)]ds)^2]^{\frac{1}{2}} \leq$$
$$\leq R(\theta,h^{\frac{1}{2}},x)\mathbb{E}[h\int_0^h (X_s^\theta - x)^2 ds]^{\frac{1}{2}} \leq R(\theta,h^{\frac{1}{2}},x)R(\theta,h^{\frac{3}{2}},x) = R(\theta,h^2,x),$$

where we have also used the first point of Lemma 1.
On the second term of (118) we act like we did on the second term of (113), using Holder inequality twice (with p big and q next to 1), (50), the polynomial growth of both a and b and (71). We get

$$\mathbb{E}[(\int_0^h a(X_s^\theta,\sigma)dW_s)(\int_0^h \partial_\mu b(X_s^\theta,\mu)ds)(1-\varphi_{h^\beta}(X_h^\theta - x))] \leq R(\theta,h^{\frac{1}{2}},x)R(\theta,h,x)R(\theta,h^{1-\epsilon},x) = R(\theta,h^{\frac{5}{2}-\epsilon},x).$$

Using on the third term of (118) Cauchy-Schwartz inequality, (50), the fact that $|\varphi|$ is bounded and the estimation of the L^p norm of G_μ given above (118) for $p=2$, we get it is upper bounded in absolute value by $R(\theta,h^2,x)$.
It follows $|I_{1,1}^\mu| \leq R(\theta,h^2,x)$, as we wanted. □

A.2.2 Proof of Proposition 9

Proof. We first of all write $\partial^2_{\mu\sigma} m$. Again, since by the homogeneity of the equation m and m_2 depend only on the difference $t_{i+1} - t_i$ we can consider, for all $h \leq \Delta_n$, $m(\mu,\sigma,x)$ as in (106). Hence, writing φ for $\varphi_{h^\beta}(X_h^\theta - x)$ (and $\varphi^{(k)}$ for $\varphi^{(k)}_{h^\beta}(X_h^\theta - x)$), we have $\partial^2_{\mu\sigma} m(\mu,\sigma,x) =$

$$\frac{\mathbb{E}[(\partial^2_{\mu\sigma}X_h^\theta)\varphi]}{\mathbb{E}[\varphi]} + \frac{2h^{-\beta}\mathbb{E}[(\partial_\mu X_h^\theta)(\partial_\sigma X_h^\theta)\varphi']}{\mathbb{E}[\varphi]} - \frac{h^{-\beta}\mathbb{E}[(\partial_\mu X_h^\theta)\varphi']\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi]}{(\mathbb{E}[\varphi])^2} + \frac{h^{-\beta}\mathbb{E}[X_h^\theta(\partial^2_{\mu\sigma}X_h^\theta)\varphi']}{\mathbb{E}[\varphi]} +$$
$$+\frac{h^{-2\beta}\mathbb{E}[X_h^\theta(\partial_\sigma X_h^\theta)(\partial_\mu X_h^\theta)\varphi'']}{\mathbb{E}[\varphi]} - \partial_\mu m\frac{h^{-\beta}\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi']}{\mathbb{E}[\varphi]} - \frac{h^{-2\beta}\mathbb{E}[X_h^\theta(\partial_\sigma X_h^\theta)\varphi']\mathbb{E}[(\partial_\mu X_h^\theta)\varphi']}{(\mathbb{E}[\varphi])^2} - \frac{h^{-\beta}m\mathbb{E}[(\partial^2_{\mu\sigma}X_h^\theta)\varphi']}{\mathbb{E}[\varphi]} + \tag{119}$$

$$-\frac{h^{-2\beta}m\mathbb{E}[(\partial_\mu X_h^\theta)(\partial_\sigma X_h^\theta)\varphi'']}{\mathbb{E}[\varphi]}+\frac{h^{-2\beta}m\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi']\mathbb{E}[(\partial_\mu X_h^\theta)\varphi']}{(\mathbb{E}[\varphi])^2}=:\sum_{j=1}^{10}I_j^n.$$

Now on I_2^n and $\sum_{j=4}^{10}I_j^n$ we use Proposition 11, where Z_h are respectively $\frac{\partial_\mu X_h^\theta}{h}\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$, $X_h^\theta\frac{\partial^2_{\mu\sigma}X_h^\theta}{h^{\frac32}}$, $X_h^\theta\frac{\partial_\mu X_h^\theta}{h}\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$, $\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$, $X_h^\theta\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$ in the first and $\frac{\partial_\mu X_h^\theta}{h}$ in the second expected value of I_7^n, $\frac{\partial^2_{\mu\sigma}X_h^\theta}{h^{\frac32}}$, $\frac{\partial_\mu X_h^\theta}{h}\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$, $\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$ in the first expected value of I_{10}^n and $\frac{\partial_\mu X_h^\theta}{h^{\frac12}}$ in the second one. Recalling moreover that $|m(\mu,\sigma,x)|\le R(\theta,1,x)$, $|\partial_\mu m(\mu,\sigma,x)|\le R(\theta,h,x)$ and the denominator gives always $R(\theta,1,x)$ it follows

$$|I_2^n+\sum_{j=4}^{10}I_j^n|\le R(\theta,h^{\frac52-\beta-\epsilon},x)+R(\theta,h^{\frac52-2\beta-\epsilon},x)+R(\theta,h^{3-\beta-\epsilon},x)+R(\theta,h^{\frac72-2\beta-\epsilon},x)=R(\theta,h^{\frac52-2\beta-\epsilon},x). \tag{120}$$

On I_1^n we use Holder inequality, the fact that $|\varphi|$ is bounded, second inequality in (96) and still the fact that the denominator is $R(\theta,1,x)$ to get

$$|I_1^n|\le R(\theta,h^{\frac32},x). \tag{121}$$

We now deal with the numerator of I_3^n, the denominator is still $R(\theta,1,x)$ as a consequence of Lemma 12. In the second expected value we apply Holder inequality, the boundedness of $|\varphi|$ and first inequality of (96) while on the first we use Proposition 11 with Z_h that this time is $\frac{\partial_\mu X_h^\theta}{h}$. We get

$$|I_3^n|\le R(\theta,h^{\frac52-\beta-\epsilon},x). \tag{122}$$

From (120), (121) and (122) it follows the first inequality in (27).
In order to prove the second inequality in (27) we compute $\partial^2_\sigma m(\mu,\sigma,x)$, getting 10 terms as in (119) in which the derivatives are always with respect to σ. We therefore say that $\partial^2_\sigma m(\mu,\sigma,x):=\sum_{j=1}^{10}\tilde I_j^n$. On $\tilde I_2^n$ and $\sum_{j=4}^{10}\tilde I_j^n$ we still use Proposition 11 taking as Z respectively $\frac{(\partial_\sigma X_h^\theta)^2}{h}$, $X_h^\theta\frac{\partial^2_\sigma X_h^\theta}{h^{\frac12}}$, $X_h^\theta\frac{(\partial_\sigma X_h^\theta)^2}{h}$, $\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$, $X_h^\theta\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$ in the first and $\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$ in the second expected value of $\tilde I_7^n$, $\frac{\partial^2_\sigma X_h^\theta}{h^{\frac12}}$, $\frac{(\partial_\sigma X_h^\theta)^2}{h}$, $\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$ in both the first and the second expected values of $\tilde I_{10}^n$. Recalling also that $|\partial_\sigma m(\mu,\sigma,x)|\le R(\theta,\Delta_{n,i},x)$ it follows

$$|\tilde I_2^n+\sum_{j=4}^{10}\tilde I_j^n|\le R(\theta,h^{2-\beta-\epsilon},x)+R(\theta,h^{\frac32-\beta-\epsilon},x)+R(\theta,h^{2-2\beta-\epsilon},x)+R(\theta,h^{\frac52-\beta-\epsilon},x)+R(\theta,h^{3-2\beta-\epsilon},x)=R(\theta,h^{\frac32-\beta-\epsilon},x). \tag{123}$$

Concerning the numerator of $\tilde I_3^n$, using Holder inequality, first inequality in (96), the boundedness of φ and Proposition 11 for $Z=\frac{\partial_\sigma X_h^\theta}{h^{\frac12}}$ it follows

$$|\tilde I_3^n|\le R(\theta,h^{2-\beta-\epsilon},x). \tag{124}$$

We now have to study $\tilde I_1^n$. We replace the dynamic (104) isolating the principal term: $\partial^2_\sigma X_h^\theta=:\int_0^h\partial^2_\sigma a(\sigma,X_s^\theta)dW_s+G_{\sigma\sigma}$.
$G_{\sigma\sigma}$ is the negligible part and, as a consequence of (105), (in which we recall that $ch^{\frac p2}(1+|x|^c)$ comes from the principal term), we already know that

$$\mathbb{E}[|G_{\sigma\sigma}|^p]\le c(1+|x|^c)(h^{2p}+h^{\frac32p}+h^p+h^{p+1})+c(h^{p-1}+h^{\frac p2-1}+1)\int_0^h\mathbb{E}[|\partial^2_\sigma X_s^\theta|^p]ds\le c(1+|x|^c)h^{\frac p2+1}. \tag{125}$$

Therefore, $\forall p\ge2$, $\mathbb{E}[|G_{\sigma\sigma}|^p]^{\frac1p}\le R(\theta,h^{\frac12+\frac1p},x)$.
We can see $\tilde I_1^n$ in the following way:

$$\tilde I_1^n=\mathbb{E}[\int_0^h\partial^2_\sigma a(\sigma,X_s^\theta)dW_s]+\mathbb{E}[(\int_0^h\partial^2_\sigma a(\sigma,X_s^\theta)dW_s)(\varphi_{h^\beta}(X_h^\theta-x)-1)]+\mathbb{E}[G_{\sigma\sigma}\varphi_{h^\beta}(X_h^\theta-x)].$$

The first expected value is zero, on the second we use Holder inequality, (50) and (71) to get it is upper bounded by $R(\theta,h^{\frac32-\epsilon},x)$. On the third term here above we use Cauchy-Schwartz inequality, the boundedness of φ and (125) for $p=2$ getting it is $R(\theta,h,x)$.
It follows second inequality in (27). Equation (28) has already been proved in Proposition 8 in [2].
Concerning the second derivatives of m_2, it is $\partial^2_{\sigma\mu}m_2=$

$$=\frac{2\mathbb{E}[(\partial_\mu X_h^\theta-\partial_\mu m)(\partial_\sigma X_h^\theta-\partial_\sigma m)\varphi]}{\mathbb{E}[\varphi]}+\frac{2\mathbb{E}[(X_h^\theta-m)(\partial^2_{\mu\sigma}X_h^\theta-\partial^2_{\mu\sigma}m)\varphi]}{\mathbb{E}[\varphi]}+\frac{2\mathbb{E}[(X_h^\theta-m)(\partial_\mu X_h^\theta-\partial_\mu m)h^{-\beta}\,\partial_\sigma X_h^\theta\,\varphi']}{\mathbb{E}[\varphi]}+$$

$$-\frac{2\mathbb{E}[(X_h^\theta-m)(\partial_\mu X_h^\theta-\partial_\mu m)h^{-\beta}\,\varphi]\mathbb{E}[\partial_\sigma X_h^\theta\,\varphi']}{(\mathbb{E}[\varphi])^2}+\frac{2\mathbb{E}[(X_h^\theta-m)(\partial_\sigma X_h^\theta-\partial_\sigma m)h^{-\beta}\,\partial_\mu X_h^\theta\,\varphi']}{\mathbb{E}[\varphi]}+$$

$$+\frac{\mathbb{E}[(X_h^\theta-m)^2(\partial_\mu X_h^\theta)h^{-2\beta}\,(\partial_\sigma X_h^\theta)\,\varphi'']}{\mathbb{E}[\varphi]}+\frac{\mathbb{E}[(X_h^\theta-m)^2(\partial^2_{\mu\sigma}X_h^\theta)h^{-\beta}\varphi']}{\mathbb{E}[\varphi]}-\frac{\mathbb{E}[(X_h^\theta-m)^2(\partial_\mu X_h^\theta)h^{-\beta}\,\varphi']\mathbb{E}[h^{-\beta}\partial_\sigma X_h^\theta\,\varphi']}{(\mathbb{E}[\varphi])^2}+ \tag{126}$$

$$-\partial_\sigma m_2\frac{\mathbb{E}[h^{-\beta}\partial_\mu X_h^\theta\,\varphi']}{\mathbb{E}[\varphi]}-m_2\frac{\mathbb{E}[h^{-2\beta}(\partial_\mu X_h^\theta)(\partial_\sigma X_h^\theta)\,\varphi'']}{\mathbb{E}[\varphi]}-m_2\frac{\mathbb{E}[h^{-\beta}(\partial^2_{\mu\sigma}X_h^\theta)\,\varphi']}{\mathbb{E}[\varphi]}+m_2\frac{\mathbb{E}[h^{-2\beta}(\partial_\mu X_h^\theta)\,\varphi']\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi']}{(\mathbb{E}[\varphi])^2}=$$

$$=:\sum_{j=1}^{12}I_j^n.$$

On I_1^n we use the first and the second point of Proposition 8 to say that both the derivatives of m are $R(\theta,h,x)$. From the boundedness of φ, Lemma 12 and the estimation of the derivatives of X gathered in Lemma 14 it follows

$$|-2\mathbb{E}[(\partial_\mu X_h^\theta-\partial_\mu m)\partial_\sigma m\varphi]|\leq R(\theta,h^2,x).$$

We now have to study

$$2\mathbb{E}[\partial_\mu X_h^\theta\partial_\sigma X_h^\theta\varphi]-2\partial_\mu m\mathbb{E}[\partial_\sigma X_h^\theta\varphi]=:I_{1,1}^n+I_{1,2}^n.$$

We start considering $I_{1,2}^n$. As we have already done after (112), we see $\partial_\sigma X_h^\theta$ as $\int_0^h\partial_\sigma a(X_s^\theta,\sigma)dW_s+G_\sigma$, hence

$$|I_{1,2}^n|\leq R(\theta,h,x)|\mathbb{E}[(\int_0^h\partial_\sigma a(X_s^\theta,\sigma)dW_s)(\varphi-1)]+\mathbb{E}[G_\sigma\varphi]|\leq R(\theta,h,x)[R(\theta,h^{\frac{3}{2}-\epsilon},x)+R(\theta,h,x)]=R(\theta,h^2,x),$$

where in the last inequality we have used on the first term Holder inequality, Burkholder - Davis - Gundy inequality and (71) while on the second we have used Cauchy - Schwartz inequality and the fact that the 2- norm of G_σ is upper bounded by a $R(\theta,h,x)$ function as a consequence of (101), (102) and (103).
Concerning $I_{1,1}^n$, replacing again $\partial_\sigma X_h^\theta$ and using the estimation on the 2-norm of G_σ as we have already done it follows

$$|I_{1,1}^n|\leq c|\mathbb{E}[(\int_0^h\partial_\sigma a(X_s^\theta,\sigma)dW_s)\partial_\mu X_h^\theta\varphi]|+R(\theta,h^2,x).$$

Now we observe we have already proved in the conclusion of Proposition 8, starting below (117), that $I_{1,1}^\mu:=2\mathbb{E}[[(\int_0^h a(X_s^\theta,\sigma)dW_s)\partial_\mu X_h^\theta\varphi]$ is such that $|I_{1,1}^\mu|\leq R(\theta,h^2,x)$. Acting exactly in the same way, considering now $\partial_\sigma a(X_s^\theta,\sigma)$ in the stochastic integral instead of $a(X_s^\theta,\sigma)$, we get that $|I_{1,1}^n|\leq R(\theta,h^2,x)$ and so, using also Lemma 12, $|I_1^n|\leq R(\theta,h^2,x)$.
Considering I_2^n, it is $I_2^n=:I_{2,1}^n+I_{2,2}^n$, where $I_{2,1}^n:=\frac{2\mathbb{E}[(X_h^\theta-m)(\partial^2_{\mu\sigma}X_h^\theta)\varphi]}{\mathbb{E}[\varphi]}$ and
$I_{2,2}^n=:(-\partial^2_{\mu\sigma}m)\frac{2\mathbb{E}[(X_h^\theta-m)\varphi]}{\mathbb{E}[\varphi]}$.
Now on $I_{2,1}^n$ we use Cauchy-Schwartz inequality, first point of Lemma 3 and (96) getting $|I_{2,1}^n|\leq R(\theta,h^2,x)$, while on $I_{2,2}^n$ we use the third point of Lemma 3 and (119) we have just proved in order to obtain $|I_{2,2}^n|\leq R(\theta,h^{\frac{5}{2}},x)$. The application of Holder inequality, Lemma 4, Lemma 14 and the first point of Proposition 8 on the numerator of I_3^n gives us

$$|I_3^n|\leq h^{-\beta}R(\theta,h^{1+\beta p},x)^{\frac{1}{p}}R(\theta,h^{\frac{3}{2}},x).$$

It is enough to take p next to 1 to get it is negligible compared to $R(\theta,h^2,x)$. We act on the first expected value of I_4^n like we did on I_2^n while on the second we use Proposition 11. It yields

$$|I_4^n|\leq h^{-\beta}R(\theta,h^{\frac{3}{2}-\epsilon},x)R(\theta,h^{\frac{3}{2}},x)=R(\theta,h^{3-\beta-\epsilon},x).$$

On I_5^n we act like we did on I_3^n, hence

$$|I_5^n|\leq h^{-\beta}R(\theta,h^{1+\beta p},x)^{\frac{1}{p}}R(\theta,h^{\frac{3}{2}-\epsilon},x),$$

that is negligible. In the same way

$$|I_6^n|\leq h^{-2\beta}R(\theta,h^{1+2\beta p},x)^{\frac{1}{p}}R(\theta,h^{\frac{3}{2}-\epsilon},x),$$

$$|I_7^n|\leq h^{-\beta}R(\theta,h^{1+2\beta p},x)^{\frac{1}{p}}R(\theta,h^{\frac{3}{2}-\epsilon},x)\quad\text{and}$$

$$|I_8^n|\leq h^{-2\beta}R(\theta,h^{1+2\beta p},x)^{\frac{1}{p}}R(\theta,h^{1-\epsilon},x)R(\theta,h^{\frac{3}{2}-\epsilon},x).$$

We recall that $|\partial_\sigma m_2(\mu,\sigma,x)| \leq R(\theta,h,x)$ and $|m_2(\mu,\sigma,x)| \leq R(\theta,h,x)$, therefore

$$|I_9^n| \leq h^{-\beta} R(\theta,h,x) R(\theta,h^{2-\epsilon},x) = R(\theta,h^{3-\beta-\epsilon},x),$$

$$|I_{10}^n| \leq h^{-2\beta} R(\theta,h,x) R(\theta,h^{\frac{5}{2}-\epsilon},x) = R(\theta,h^{\frac{7}{2}-2\beta-\epsilon},x),$$

$$|I_{11}^n| \leq h^{-\beta} R(\theta,h,x) R(\theta,h^{\frac{5}{2}-\epsilon},x) = R(\theta,h^{\frac{7}{2}-\beta-\epsilon},x),$$

$$|I_{12}^n| \leq h^{-2\beta} R(\theta,h,x) R(\theta,h^{2-\epsilon},x) R(\theta,h^{\frac{3}{2}-\epsilon},x) = R(\theta,h^{\frac{9}{2}-2\beta-\epsilon},x).$$

First inequality in (29) follows. In order to show the second one we should compute $\partial_\mu^2 m_2(\mu,\sigma,x)$. Since it is exactly like $\partial_{\mu\sigma}^2 m_2(\mu,\sigma,x)$ but with all the derivatives with respect to μ, we still refer to (126) and we will write $\partial_\mu^2 m_2(\mu,\sigma,x) =: \sum_{j=1}^{12} \tilde{I}_j^n$.
On $\tilde{I}_1^n$, $\sum_{j=3}^{12} \tilde{I}_j^n$ we act exactly like we did here above, getting

$$|\tilde{I}_1^n + \sum_{j=3}^{12} \tilde{I}_j^n| \leq R(\theta,h^2,x) + h^{-\beta} R(\theta,h^{1+\beta p},x)^{\frac{1}{p}} R(\theta,h^2,x) + h^{-\beta} R(\theta,h^{\frac{3}{2}},x) R(\theta,h^{2-\epsilon},x) + h^{-\beta} R(\theta,h^{1+\beta p},x)^{\frac{1}{p}} R(\theta,h^2,x) +$$

$$+ h^{-2\beta} R(\theta,h^{1+2\beta p},x)^{\frac{1}{p}} R(\theta,h^2,x) + h^{-\beta} R(\theta,h^{1+2\beta p},x)^{\frac{1}{p}} R(\theta,h,x) + h^{-2\beta} R(\theta,h^{1+2\beta p},x)^{\frac{1}{p}} R(\theta,h,x) R(\theta,h^{2-\epsilon},x) +$$
$$+ h^{-\beta} R(\theta,h^2,x) R(\theta,h^{2-\epsilon},x) + h^{-2\beta} R(\theta,h,x) R(\theta,h^{3-\epsilon},x) + h^{-\beta} R(\theta,h,x) R(\theta,h^{2-\epsilon},x) + h^{-2\beta} R(\theta,h,x) R(\theta,h^{4-\epsilon},x),$$

that is a $R(\theta,h^2,x)$ function as a consequence of the fact that, choosing p next to 1, all terms are negligible compared to the first one.
We now deal with $\tilde{I}_2^n$. We still need to split it in $\tilde{I}_{2,1}^n$ and $\tilde{I}_{2,2}^n$, where $\tilde{I}_{2,1}^n := \frac{2\mathbb{E}[(X_h^\theta - m)(\partial_\mu^2 X_h^\theta)\varphi]}{\mathbb{E}[\varphi]}$ and $\tilde{I}_{2,2}^n =: (-\partial_\mu^2 m)\frac{2\mathbb{E}[(X_h^\theta - m)\varphi]}{\mathbb{E}[\varphi]}$.
Using on $\tilde{I}_{2,2}^n$ Lemma 3 and (28) we obtain $|\tilde{I}_{2,2}^n| \leq R(\theta,h^2,x)$.
In order to estimate $\tilde{I}_{2,1}^n$ we isolate the principal term in the dynamic of $\partial_\mu^2 X_h^\theta$, getting $\partial_\mu^2 X_h^\theta = \int_0^h \partial_\mu^2 b(X_s^\theta,\mu) ds + G_{\mu\mu}$, where $G_{\mu\mu}$ is the negligible part and it is such that, for $p \geq 2$,

$$\mathbb{E}[|G_{\mu\mu}|^p] \leq c(1+|x|^c)(h^{2p} + h^{p+1}) = R(\theta,h^{p+1},x), \tag{127}$$

as showed in the proof of Lemma 9 in [2].
Replacing in the definition of $\tilde{I}_{2,1}^n$ we have

$$|\tilde{I}_{2,1}^n| \leq |\frac{2\mathbb{E}[G_{\mu\mu}(X_h^\theta - m)\varphi]}{\mathbb{E}[\varphi]}| + |\frac{2\mathbb{E}[\int_0^h \partial_\mu^2 b(X_s^\theta,\mu) ds (X_h^\theta - m)\varphi]}{\mathbb{E}[\varphi]}|.$$

Now on the first term here above we use Cauchy-Schwartz inequality, the first point of Lemma 3 and (127) obtaining it is upper bounded by $R(\theta,h^{\frac{3}{2}},x)R(\theta,h^{\frac{1}{2}},x) = R(\theta,h^2,x)$, while the second one is upper bounded by

$$ch\partial_\mu^2 b(x,\mu)|\mathbb{E}[(X_h^\theta - m)\varphi]| + c|\mathbb{E}[\int_0^h [\partial_\mu^2 b(X_s^\theta,\mu) - \partial_\mu^2 b(x,\mu)] ds (X_h^\theta - m)\varphi]| \leq$$

$$\leq hR(\theta,h,x) + R(\theta,h^{\frac{3}{2}},x)R(\theta,h^{\frac{1}{2}},x) = R(\theta,h^2,x),$$

where we have also used Lemma 12 in order to say that the denominator is lowed bounded by $1+R(\theta,h,x)$, Cauchy-Schwartz inequality, the first point of Lemma 3 and an estimation analogous to (82) for the increments of the derivative of b. It follows $|\tilde{I}_{2,1}^n| \leq R(\theta,h^2,x)$.
Second inequality in (29) follows.

We now want to show (30). To do it, we need to compute $\partial_\sigma^2 m_2(\mu,\sigma,x)$. Again, we still refer to (126) considering all the derivatives with respect to σ, writing $\partial_\sigma^2 m_2(\mu,\sigma,x) =: \sum_{j=1}^{12} \hat{I}_j^n$. From $\hat{I}_3^n$ to $\hat{I}_{12}^n$ all the terms are negligible. Indeed, acting like we did on both $\sum_{j=3}^{12} I_j^n$ and $\sum_{j=3}^{12} \tilde{I}_j^n$ we have

$$|\sum_{j=3}^{12} \hat{I}_j^n| \leq h^{-\beta} R(\theta,h^{1+\beta p},x)^{\frac{1}{p}} R(\theta,h,x) + h^{-\beta} R(\theta,h^{1+\beta p},x)^{\frac{1}{p}} R(\theta,h^{2-\epsilon},x) + h^{-\beta} R(\theta,h^{1+\beta p},x)^{\frac{1}{p}} R(\theta,h,x) +$$

$$+ h^{-2\beta} R(\theta,h^{1+2\beta p},x)^{\frac{1}{p}} R(\theta,h,x) + h^{-\beta} R(\theta,h^{1+2\beta p},x)^{\frac{1}{p}} R(\theta,h^{\frac{1}{2}},x) + h^{-2\beta} R(\theta,h^{1+2\beta p},x)^{\frac{1}{p}} R(\theta,h^{\frac{1}{2}},x) R(\theta,h^{\frac{3}{2}-\epsilon},x) +$$
$$+ h^{-\beta} R(\theta,h,x) R(\theta,h^{\frac{3}{2}-\epsilon},x) + h^{-2\beta} R(\theta,h,x) R(\theta,h^{2-\epsilon},x) + h^{-\beta} R(\theta,h,x) R(\theta,h^{\frac{3}{2}-\epsilon},x) + h^{-2\beta} R(\theta,h,x) R(\theta,h^{3-2\epsilon},x),$$

that is always negligible compared to $R(\theta, h^{\frac{3}{2},x})$ taking p next to 1. We now consider $\hat{I}_1^n$. Its numerator is

$$\mathbb{E}[(\partial_\sigma X_h^\theta)^2\varphi] + (\partial_\sigma m)^2\mathbb{E}[\varphi] - 2(\partial_\sigma m)\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi] =: \hat{I}_{1,1}^n + \hat{I}_{1,2}^n + \hat{I}_{1,3}^n.$$

We start considering $\hat{I}_{1,1}^n$, in which we replace the dynamic of $\partial_\sigma X_h^\theta := \int_0^h \partial_\sigma a(\sigma, X_s^\theta)dW_s + G_\sigma$, where we have already seen that, for $p \geq 2$, $\mathbb{E}[|G_\sigma|^p] \leq c(h^{\frac{3p}{2}} + h^p + h^{\frac{p}{2}+1})(1 + |x|^c) = R(\theta, h^{\frac{p}{2}+1}, x)$. We therefore can say that, taking $p = 2$ and using (50), we have

$$\hat{I}_{1,1}^n = \mathbb{E}[(\int_0^h \partial_\sigma a(\sigma, X_s^\theta)dW_s)^2\varphi] + R(\theta, h^{\frac{3}{2}}, x) + R(\theta, h^2, x).$$

The first term here above can be seen as

$$\mathbb{E}[(\int_0^h \partial_\sigma a(\sigma, X_s^\theta)dW_s)^2(\varphi - 1)] + \mathbb{E}[\int_0^h [(\partial_\sigma a)^2(\sigma, X_s^\theta) - (\partial_\sigma a)^2(\sigma, x)]ds] + h(\partial_\sigma a)^2(\sigma, x).$$

Acting like we did in (73) - (75), we have

$$|\mathbb{E}[(\int_0^h \partial_\sigma a(\sigma, X_s^\theta)dW_s)^2(\varphi - 1)] + \mathbb{E}[\int_0^h [(\partial_\sigma a(\sigma, X_s^\theta))^2 - (\partial_\sigma a(\sigma, x))^2]ds]| \leq R(\theta, h^{\frac{3}{2}}, x)$$

and so we get

$$\hat{I}_{1,1}^n = h(\partial_\sigma a(\sigma, x))^2 + R(\theta, h^{\frac{3}{2}}, x).$$

On $\hat{I}_{1,2}^n$ we use the second point of Proposition 8 and Lemma 12 to get $|\hat{I}_{1,2}^n| \leq R(\theta, h^2, x)$. In the same way, from second point of Proposition 8, the boundedness of φ and (96) we get $|\hat{I}_{1,3}^n| \leq R(\theta, h^{\frac{3}{2}}, x)$. It follows $\hat{I}_1^n = h(\partial_\sigma a(\sigma, x))^2 + R(\theta, h^{\frac{3}{2}}, x)$.
We now study $\hat{I}_2^n$, that the denominator is still $1 + R(\theta, h, x)$ as a consequence of Lemma 12 while the numerator can be seen as

$$2\mathbb{E}[(X_h^\theta - m)\partial_\sigma^2 X_h^\theta\varphi] - 2\partial_\sigma^2 m\mathbb{E}[(X_h^\theta - m)\varphi] =: \hat{I}_{2,1}^n + \hat{I}_{2,2}^n.$$

In order to deal with $\hat{I}_{2,1}^n$ we consider the reformulation (72), so we have $2\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma)dW_s)(\partial_\sigma^2 X_h^\theta)\varphi] + 2\mathbb{E}[(B_h)(\partial_\sigma^2 X_h^\theta)\varphi]$.
The second, as a consequence of Holder inequality, the definition of B_h, Lemma 5 used to estimate the jumps and the first inequality in (96) is upper bounded by $R(\theta, h^{(1+\beta p)\wedge p}, x)^{\frac{1}{p}} R(\theta, h^{\frac{1}{2}}, x)$. Taking p next to 1 it follows its order is $h^{\frac{3}{2}}$.
In order to study $\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma)dW_s)(\partial_\sigma^2 X_h^\theta)\varphi]$ we introduce once again the notation used above (125), for which $\partial_\sigma^2 X_h^\theta = \int_0^h \partial_\sigma^2 a(\sigma, X_s^\theta)dW_s + G_{\sigma\sigma}$. Hence, we have

$$\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma)dW_s)(\int_0^h \partial_\sigma^2 a(\sigma, X_s^\theta)dW_s)\varphi] + \mathbb{E}[(\int_0^h a(X_s^\theta, \sigma)dW_s)(G_{\sigma\sigma})\varphi].$$

The second term here above is, using Cauchy-Schwartz inequality, (50), the boundedness of φ and (125), just $R(\theta, h^{\frac{3}{2}}, x)$ while the first one is

$$\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma)dW_s)(\int_0^h \partial_\sigma^2 a(\sigma, X_s^\theta)dW_s)(\varphi - 1)] + \mathbb{E}[\int_0^h a(X_s^\theta, \sigma)\partial_\sigma^2 a(\sigma, X_s^\theta)ds] = R(\theta, h^2, x) +$$

$$+\mathbb{E}[\int_0^h [a(X_s^\theta, \sigma) - a(x, \sigma)]\partial_\sigma^2 a(\sigma, X_s^\theta)ds] + a(x, \sigma)\mathbb{E}[\int_0^h [\partial_\sigma^2 a(\sigma, X_s^\theta) - \partial_\sigma^2 a(\sigma, x)]ds)] + ha(x, \sigma)\partial_\sigma^2 a(\sigma, x),$$

where we have also used Holder inequality, (50) and (71). Now we observe that

$$|\mathbb{E}[\int_0^h [a(X_s^\theta, \sigma) - a(x, \sigma)]\partial_\sigma^2 a(\sigma, X_s^\theta)ds]| \leq c\int_0^h \mathbb{E}[|X_s^\theta - x|^2]^{\frac{1}{2}}\mathbb{E}[|\partial_\sigma^2 a(\sigma, X_s^\theta)|^2]^{\frac{1}{2}}ds \leq \int_0^h R(\theta, h^{\frac{1}{2}}, x) = R(\theta, h^{\frac{3}{2}}, x).$$

Acting in the same way on $a(x, \sigma)\mathbb{E}[\int_0^h [\partial_\sigma^2 a(\sigma, X_s^\theta) - \partial_\sigma^2 a(\sigma, x)]ds)]$ it follows that

$$\hat{I}_{2,1}^n = ha(x, \sigma)\partial_\sigma^2 a(\sigma, x) + R(\theta, h^{\frac{3}{2}}, x).$$

On $\hat{I}_{2,2}^n$ we use the second inequality in (27) and the third point of Lemma 3 and so we obtain

$$|\hat{I}_{2,2}^n| \leq R(\theta, h^2, x),$$

the result follows. □

A.2.3 Proof of Proposition 10

Proof. We define $\partial_\mu^2 m_2(\mu,\sigma,x)$ as $\sum_{j=1}^{12}\tilde{I}_j^n =: \sum_{j=1}^{12}\frac{N_{\tilde{I}_j^n}}{D_{\tilde{I}_j^n}}$ and we recall that, as stated in Proposition 9, $|\tilde{I}_j^n| \leq R(\theta,\Delta_{n,i}^2,x)$. We can therefore see its derivative with respect to μ in the following way:

$$\partial_\mu^3 m_2(\mu,\sigma,x) = \sum_{j=1}^{12}\left(\frac{\partial_\mu N_{\tilde{I}_j^n}}{D_{\tilde{I}_j^n}} - \tilde{I}_j^n\frac{\partial_\mu D_{\tilde{I}_j^n}}{D_{\tilde{I}_j^n}}\right). \tag{128}$$

We start considering the second term here above: we remind that $D_{\tilde{I}_j^n}$ can be $\mathbb{E}[\varphi]$ or $(\mathbb{E}[\varphi])^2$. In both cases its derivative is, using Proposition 11, a $R(\theta,h^{2-\beta-\epsilon},x)$ function; which makes the second term of (128) upper bounded by $\sum_{j=1}^{12}|\tilde{I}_j^n|R(\theta,h^{2-\beta-\epsilon},x) \leq R(\theta,h^{4-\beta-\epsilon},x)$, as a consequence of the second inequality of (29).
Concerning the first term of (128), we know that $D_{\tilde{I}_j^n}$ is a $R(\theta,1,x)$ function because of Lemma 12, while the magnitude of the derivative of the numerator does not get bigger since the order of X_h^θ and m remains the same by deriving them once more (as gathered in the second inequality of (98) and in Remark 10 of [2]) and the fact that, by deriving φ, there comes out $h^{-\beta}\varphi'\partial_\mu X_h^\theta$. From the first inequality in (95) such terms are negligible compared to the ones we have already studied.
The first term of (128) is therefore still a $R(\theta,h^2,x)$ function.
The same reasoning applies to the study of the other seven third derivatives. □

A.3 Development of $m_2(\mu,\sigma,x)$

In order to prove our main results we need a development of m_2, that we stated in Ad and we find, under the choice of a particular oscillating function φ, in Propositions 2, 3 and 4.
The main tools is the iteration of the Dynkin's formula that provides us the following expansion for every function $f\colon \mathbb{R}\to\mathbb{R}$ such that f is in $C^{2(k+1)}$:

$$\mathbb{E}[f(X_{t_{i+1}}^\theta)|X_{t_i}^\theta = x] = \sum_{j=0}^{k}\frac{\Delta_{n,i}^j}{j!}A^j f(x) + \int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}\dots\int_{t_i}^{u_k}\mathbb{E}[A^{k+1}f(X_{u_{k+1}}^\theta)|X_{t_i}^\theta = x]\,du_{k+1}\dots du_2\,du_1 \tag{129}$$

where A denotes the generator of the diffusion, setting $A^0 = Id$. A is the sum of the continuous and discrete part: $A := \bar{A}_c + A_d$, with

$$\bar{A}_c f(x) = \frac{1}{2}a^2(x,\sigma)f''(x) + \bar{b}(x,\mu)f'(x);$$

$\bar{b}(x,\mu) = b(x,\mu) - \int_{\mathbb{R}} z\gamma(x)F(z)dz$ and

$$A_d f(x) = \int_{\mathbb{R}}\left(f(x+\gamma(x)z) - f(x)\right)F(z)dz.$$

A.3.1 Proof of Proposition 2.

Proof. We first of all observe that, by the definition of m_2 and m it is

$$m_2(\mu,\sigma,x) = \frac{\mathbb{E}[(X_{t_{i+1}}^\theta)^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]} - (m(\mu,\sigma,x))^2.$$

It has already been showed in Proposition 2 in [2] that

$$m(\mu,\sigma,x) = x + \sum_{k=1}^{\lfloor\beta(M+2)\rfloor}A_{K_1}^{(k)}(x)\frac{\Delta_{n,i}^k}{k!} + R(\theta,\Delta_{n,i}^{\beta(M+2)},x),$$

where $A_{K_1}^{(k)}(x) = \bar{A}_c(h_1)(x)$, with $h_1(y) = (y-x)$.
Acting like we did in Proposition 2 of [2], we want to find a development for

$$\frac{\mathbb{E}[(X_{t_{i+1}}^\theta)^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]} =: \frac{n_{\Delta_{n,i}}(x)}{d_{\Delta_{n,i}}(x)}.$$

First we focus on the expression of n. We define the set of rest functions $\mathcal{R}^p$ in the following way:

$$\mathcal{R}^p := \left\{ r(x,y,\Delta_{n,i}^p,\theta) \text{ s. t. } \forall l\geq 0, \forall l'\in\{0,1\}\ , \exists c, |\frac{\partial^l\,\partial^{l'}}{\partial y^l\,\partial\vartheta^{l'}}r(x,y,\Delta_{n,i}^p,\theta)| \leq c(1+|x|^c+|y|^c)\Delta_{n,i}^p, \text{ for } \vartheta=\mu,\ \vartheta=\sigma \right\}.$$

It is worth noting that, if $r \in \mathcal{R}^p$, then both $\bar{A}_c r$ and $A_d r$ are in $\mathcal{R}^p$, where the integral-differential operators $\bar{A}_c$ and A_d are applied with respect to the second variable y (see details in proof of Proposition 2 in [2]).
We also introduce the set $\tilde{\mathcal{F}}^p$:

$$\tilde{\mathcal{F}}^p := \left\{ \tilde{g}(y,\theta) \; s.t. \; \tilde{g}(y,\theta) = \sum_{k=0}^{p} \varphi^{(k)}((y-x)\Delta_{n,i}^{-\beta})\Delta_{n,i}^{-k\beta}(\sum_{j=0}^{k} h_{k,j}(x,y,\theta)\Delta_{n,i}^{\beta j}) \right\}$$

where, $\forall k,j$, $\forall l \geq 0$, $\forall l' \in \{0,1\}$, $\exists c$ such that, for $\vartheta = \mu$ and $\vartheta = \sigma$, $\sup_{\theta\in\Theta} |\frac{\partial^l}{\partial y^l}\frac{\partial^{l'}}{\partial \vartheta^{l'}} h_{k,j}(x,y,\theta)| \leq c(1+|x|^c+|y|^c)$, with c that depends on k,j,l and l'. By the same proof as in Proposition 2 of [2] it is possible to prove that, if $\tilde{g} \in \tilde{\mathcal{F}}^p$ then $\bar{A}_c\tilde{g} \in \tilde{\mathcal{F}}^{p+2}$ and, for all $\tilde{g} \in \tilde{\mathcal{F}}^p$,

$$A_d\tilde{g}(y,\theta) = -\lambda\tilde{g}(y,\theta) + r(x,y,\Delta_{n,i}^{\beta(M+2-p)},\theta). \tag{130}$$

It has also been proved that, as a consequence of the equation here above, the following relation holds:

$$A_{i_N} \circ ... \circ A_{i_1}\tilde{g}(y) = \bar{A}_c^{l(i_1,...,i_N)}\tilde{g}(y)(-\lambda)^{N-l(i_1,...,i_N)} + r(x,y,\Delta_{n,i}^{\beta(M+2)-2\beta l(i_1,...,i_N)}),$$

for $\tilde{g} \in \tilde{\mathcal{F}}_0$ and with $l(i_1,...,i_N)$ the number of c in $\{i_1,...,i_N\}$, the iterations considered.
We observe that $\bar{g}(y) := y^2\varphi((y-x)\Delta_{n,i}^{-\beta})$ belongs to $\tilde{\mathcal{F}}_0$. To find its development through Dynkin's formula we can act exactly like we did in Proposition 2 of [2].
We get that the principal term in the development of the numerator is

$$\sum_{l=0}^{N} \frac{\Delta_{n,i}^l}{l!} A_{K_2}^{(l)}(x) \sum_{k'=0}^{N-l} \frac{\Delta_{n,i}^{k'}(-\lambda)^{k'}}{k'!} + r(x,x,\Delta_{n,i}^{\beta(M+2)},\theta), \tag{131}$$

with $A_{K_2}(x) = \bar{A}_c(h_2)(x)$ for $h_2(y) = y^2$ and $r(x,y,\Delta_{n,i}^{\beta(M+2)},\theta) \in \mathcal{R}^{\beta(M+2)}$. The integral rest term in the Dynkin formula is

$$|\int_{t_i}^{t_{i+1}} \int_{t_i}^{u_1} ... \int_{t_i}^{u_N} \mathbb{E}[A^{N+1}h_2(X_{u_{N+1}})|X_{t_i} = x]du_{N+1}...du_2du_1| \leq R(\theta,\Delta_{n,i}^{(1-2\beta)(N+1)},x). \tag{132}$$

Using (131) and (132) we have the following development:

$$n_{\Delta_{n,i}}(x) = \sum_{l=0}^{N} \frac{\Delta_{n,i}^l}{l!} A_{K_2}^{(l)}(x) \sum_{k'=0}^{N-l} \frac{\Delta_{n,i}^{k'}(-\lambda)^{k'}}{k'!} + R(\theta,\Delta_{n,i}^{\beta(M+2)},x) + R(\theta,\Delta_{n,i}^{(1-2\beta)(N+1)},x). \tag{133}$$

If $(N+1)(1-2\beta) \geq \beta(M+2)$, it entails

$$n_{\Delta_{n,i}}(x) = \sum_{l=0}^{\lfloor\beta(M+2)\rfloor} \frac{\Delta_{n,i}^l}{l!} A_{K_2}^{(l)}(x) \sum_{k'=0}^{\lfloor\beta(M+2)\rfloor} \frac{\Delta_{n,i}^{k'}(-\lambda)^{k'}}{k'!} + R(\theta,\Delta_{n,i}^{\beta(M+2)},x). \tag{134}$$

To get the control (12) on the derivatives of (134), we show that one can differentiate with respect to the parameters the remainder term.
We remark that, as $r(x,y,\Delta_{n,i}^{\beta(M+2)},\theta)$ is in $\mathcal{R}^{\beta(M+2)}$ and because of the definition of such a set, we clearly have that for $\vartheta = \mu$ and $\vartheta = \sigma$, the function $\partial_\vartheta r(x,y,\Delta_{n,i}^{\beta(M+2)},\theta)$ is still a $R(\theta,\Delta_{n,i}^{\beta(M+2)},x)$ function.
Concerning the derivatives of the integral rest, we have that, since $A^{N+1}\bar{g} \in \tilde{\mathcal{F}}^{2(N+1)}$,

$$A^{N+1}\bar{g}(y) = \sum_{k=0}^{2N} \varphi^{(k)}((y-x)\Delta_{n,i}^{-\beta})\Delta_{n,i}^{-\beta k}(\sum_{j=0}^{k} \tilde{h}_{k,j}(x,y,\theta)\Delta_{n,i}^{\beta j});$$

where $\tilde{h}_{k,j}$ are polynomial functions of a, b and their derivatives. Then, for $\vartheta = \mu$ and $\vartheta = \sigma$ it is

$$\partial_\vartheta \mathbb{E}[A^{N+1}\bar{g}(X_{u_{N+1}})|X_{t_i} = x] = \mathbb{E}[\partial_\vartheta A^{N+1}\bar{g}(X_{u_{N+1}}) + \partial_X A^{N+1}\bar{g}(X_{u_{N+1}})\partial_\vartheta X_{u_{N+1}}|X_{t_i} = x].$$

We use an upper bound on the conditional moments of the derivative of X with respect to both the parameters (see Lemma 14 in the Appendix) and we act as for (132) to get a control of the integral rest. From the computation of $\partial_X A^{N+1}\bar{g}(X_{u_{N+1}})$ it shows up an extra $\Delta_{n,i}^{-\beta}$ but we can always choose N such that the derivatives of the integral rest remain negligible compared to $R(\theta,\Delta_{n,i}^{\beta(M+2)},x)$.

We now consider the denominator $d_{\Delta_{n,i}}(x)$: since it is exactly like it was in the development of m, we have already proved in Proposition 2 of [2] that its expansion is

$$d_{\Delta_{n,i}}(x) = \sum_{k=0}^{\lfloor \beta(M+2) \rfloor} \frac{\Delta_{n,i}^k}{k!}(-\lambda)^k + r(x,x,\Delta_{n,i}^{\beta(M+2)}), \tag{135}$$

where λ is the intensity of jumps defined in the fourth point of Assumption 4 and r is a rest function which belongs to $\mathcal{R}^{\beta(M+2)}$.
Acting exactly as we have done on the numerator we get that the derivatives of $r(x,x,\Delta_{n,i}^{\beta(M+2)})$ are still $R(\theta,\Delta_{n,i}^{\beta(M+2)},x)$ functions and that the derivatives of the integral rest remain negligible compared to $R(\theta,\Delta_{n,i}^{\beta(M+2)},x)$.
From the expansion of $d_{\Delta_{n,i}}(x)$ we can say that there exists $k_0 > 0$ such that for $|x| \leq \Delta_{n,i}^{-k_0}$, $d_{\Delta_{n,i}}(x) \geq \frac{1}{2}$ $\forall n, i \leq n$: we are avoiding the possibility that the denominator is in the neighborhood of 0. Hence, for $|x| \leq \Delta_{n,i}^{k_0}$, the development of m_2 is

$$\frac{n_{\Delta_{n,i}}(x)}{d_{\Delta_{n,i}}(x)} - (m(\mu,\sigma,x))^2 = \sum_{l=0}^{\lfloor \beta(M+2) \rfloor} \frac{\Delta_{n,i}^l}{l!} A_{K_2}^{(l)}(x) - (x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_{K_1}^{(k)}(x)\frac{\Delta_{n,i}^k}{k!})^2 + r(x,x,\Delta_{n,i}^{\beta(M+2)},\theta). \tag{136}$$

The expansion (11) follows after remarking that $A_{K_2}^{(0)}(x) = 0$.
Moreover, since the rest term in the development of m comes from the same place as the rest in the fraction just studied, also its derivatives with respect to μ and σ remain $R(\theta,\Delta_{n,i}^{\beta(M+2)},x)$ functions. The result follows. □

We now prove Proposition 4, Proposition 3 is a consequence of it.

A.3.2 Proof of Proposition 4

We recall that

$$m_2(\mu,\sigma,x) = \frac{\mathbb{E}[(X_{t_{i+1}}^\theta - m(\mu,\sigma,X_{t_i}))^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]} := \frac{\tilde{n}_{\Delta_{n,i}}(x)}{\tilde{d}_{\Delta_{n,i}}(x)}.$$

The expansion (15) in Proposition 4 is a consequence of the following two expansions for $\tilde{n}_{\Delta_{n,i}}(x)$ and $\tilde{d}_{\Delta_{n,i}}(x)$,

$$\begin{aligned}\tilde{n}_{\Delta_{n,i}}(x) = \Delta_{n,i}a^2(x,\sigma) + \frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)}\int_{\mathbb{R}} v^2\varphi(v)F(\frac{v\Delta_{n,i}^\beta}{\gamma(x)})dv + \Delta_{n,i}^2[3\bar{b}^2(x,\mu) + h_2(x,\theta) - \lambda a^2(x,\sigma)] \\ + \frac{\Delta_{n,i}^{2+\beta}a^2(x,\sigma)}{2\gamma(x)}\int_{u:|u|\leq 2}[2\varphi(u) + u\varphi'(u) + u^2\varphi''(u)]F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)})du + R(\theta,\Delta_{n,i}^{(3-2\beta)\wedge(2+\beta)},x),\end{aligned} \tag{137}$$

$$\tilde{d}_{\Delta_{n,i}}(x) = 1 - \Delta_{n,i}\lambda + \frac{\Delta_{n,i}^{1+\beta}}{\gamma(x)}\int_{u:|u|\leq 2}\varphi(u)F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)})du + R(\theta,\Delta_{n,i}^{2-2\beta},x). \tag{138}$$

The fact that the order of the remainder term in (15) is unchanged by derivation with respect to the parameters μ and σ, will be a consequence of a similar property for the remainder terms appearing in (137)–(138).

Proof of (137). In order to develop the numerator $\tilde{n}_{\Delta_{n,i}}(x)$, we define the function $h_{i,n}(y) := (y - m(\mu,\sigma,x))^2\varphi_{\Delta_{n,i}^\beta}(y-x)$ and we use Dynkin formula (129) on it, up to third order. It becomes

$$\mathbb{E}[h_{i,n}(X_{t_{i+1}}^\theta)|X_{t_i}^\theta = x] = h_{i,n}(x) + \Delta_{n,i}Ah_{i,n}(x) + \frac{1}{2}\Delta_{n,i}^2A^2h_{i,n}(x) + \frac{1}{6}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}\int_{t_i}^{u_2}\mathbb{E}[A^3h_{i,n}(X_{u_3}^\theta)|X_{t_i}^\theta = x]du_3du_2du_1. \tag{139}$$

We now successively study the contribution of each term in the Dynkin's development.

By the definition of $h_{i,n}$, we have $h_{i,n}(x) = (x - m(\mu,\sigma,x))^2$.

We recall that $Ah_{i,n}(x) = \bar{A}_ch_{i,n}(x) + A_dh_{i,n}(x)$, where $\bar{A}_ch_{i,n}(x) = \frac{1}{2}a^2(x,\sigma)h_{i,n}''(x) + \bar{b}(x,\mu)h_{i,n}'(x)$ and

$$A_dh_{i,n}(x) = \int_{\mathbb{R}}[h_{i,n}(x + z\gamma(x)) - h_{i,n}(x)]F(z)dz = \int_{\mathbb{R}}(x + z\gamma(x) - m(\mu,\sigma,x))^2\varphi_{\Delta_{n,i}^\beta}(z\gamma(x))F(z)dz +$$

$$-(x-m(\mu,\sigma,x))^2\int_{\mathbb{R}}\varphi_{\Delta_{n,i}^\beta}(0)F(z)dz=(x-m(\mu,\sigma,x))^2\int_{\mathbb{R}}[\varphi_{\Delta_{n,i}^\beta}(z\gamma(x))-\varphi_{\Delta_{n,i}^\beta}(0)]F(z)dz+ \quad (140)$$

$$+\gamma^2(x)\int_{\mathbb{R}}z^2\varphi_{\Delta_{n,i}^\beta}(z\gamma(x))F(z)dz+2(x-m(\mu,\sigma,x))\gamma(x)\int_{\mathbb{R}}z\varphi_{\Delta_{n,i}^\beta}(z\gamma(x))F(z)dz.$$

Using now the boundedness of φ, the fact that $\int_{\mathbb{R}}F(z)dz=\lambda$ and the development of m (31), we have that the first term here above is a $R(\theta,\Delta_{n,i}^2,x)$ function and, from Proposition 8, also its derivatives with respect to both the parameters are $R(\theta,\Delta_{n,i}^2,x)$ functions.

On the second term of (140) we apply the change of variable $v:=\frac{\gamma(x)z}{\Delta_{n,i}^\beta}$, getting $\frac{\Delta_{n,i}^{3\beta}}{\gamma(x)}\int_{\mathbb{R}}v^2\varphi(v)F(\frac{v\Delta_{n,i}^\beta}{\gamma(x)})dv$.
On the third we use the definition of φ for which $\varphi(\zeta)=0$ for ζ such that $|\zeta|\geq 2$ and again the development of m to get it is upper bounded by

$$|c(x-m(\mu,\sigma,x))\gamma(x)\int_{z:|z|\leq\frac{2\Delta_{n,i}^\beta}{\gamma(x)}}zF(z)dz|\leq R(\theta,\Delta_{n,i}^{1+\beta},x). \quad (141)$$

Again, we can calculate the derivatives with respect to ϑ for both $\vartheta=\mu$ and $\vartheta=\sigma$, getting a term that is still a $R(\theta,\Delta_{n,i}^{1+\beta},x)$ function by Proposition 8.
It follows

$$A_dh_{i,n}(x)=R(\theta,\Delta_{n,i}^2,x)+\Delta_{n,i}^{2\beta}\int_{\mathbb{R}}v^2\varphi(v)F(\frac{v\Delta_{n,i}^\beta}{\gamma(x)})dv+R(\theta,\Delta_{n,i}^{1+\beta},x).$$

In order to compute $\bar{A}_ch_{i,n}(x)$ we need the derivatives of $h_{i,n}$, they are

$$h'_{i,n}(y)=2(y-m(\mu,\sigma,x))\varphi_{\Delta_{n,i}^\beta}(y-x)+(y-m(\mu,\sigma,x))^2\Delta_{n,i}^{-\beta}\varphi'_{\Delta_{n,i}^\beta}(y-x),$$

$$h''_{i,n}(y)=2\varphi_{\Delta_{n,i}^\beta}(y-x)+4(y-m(\mu,\sigma,x))\Delta_{n,i}^{-\beta}\varphi'_{\Delta_{n,i}^\beta}(y-x)+(y-m(\mu,\sigma,x))^2\Delta_{n,i}^{-2\beta}\varphi''_{\Delta_{n,i}^\beta}(y-x), \quad (142)$$

$$h^{(3)}_{i,n}(y)=6\Delta_{n,i}^{-\beta}\varphi'_{\Delta_{n,i}^\beta}(y-x)+6(y-m(\mu,\sigma,x))\Delta_{n,i}^{-2\beta}\varphi''_{\Delta_{n,i}^\beta}(y-x)+(y-m(\mu,\sigma,x))^2\Delta_{n,i}^{-3\beta}\varphi^{(3)}_{\Delta_{n,i}^\beta}(y-x),$$

$$h^{(4)}_{i,n}(y)=12\Delta_{n,i}^{-2\beta}\varphi''_{\Delta_{n,i}^\beta}(y-x)+8(y-m(\mu,\sigma,x))\Delta_{n,i}^{-3\beta}\varphi^{(3)}_{\Delta_{n,i}^\beta}(y-x)+(y-m(\mu,\sigma,x))^2\Delta_{n,i}^{-4\beta}\varphi^{(4)}_{\Delta_{n,i}^\beta}(y-x);$$

we have calculated the derivatives up to the fourth because they will be useful in the sequel.
Replacing the first two derivatives, calculated in x, it follows

$$\bar{A}_ch_{i,n}(x)=a^2(x,\sigma)+2(x-m(\mu,\sigma,x))\bar{b}(x,\mu). \quad (143)$$

Therefore we have

$$\mathbb{E}_i[h_{i,n}(X^\theta_{t_{i+1}})]=(x-m(\mu,\sigma,x))^2+\Delta_{n,i}a^2(x,\sigma)+\frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)}\int_{\mathbb{R}}v^2\varphi(v)F(\frac{v\Delta_{n,i}^\beta}{\gamma(x)})dv+2\Delta_{n,i}(x-m(\mu,\sigma,x))\bar{b}(x,\mu)+ \quad (144)$$

$$+R(\theta,\Delta_{n,i}^{2+\beta},x)+\frac{1}{2}\Delta_{n,i}^2A^2h_{i,n}(x)+\frac{1}{6}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}\int_{t_i}^{u_2}\mathbb{E}[A^3h_{i,n}(X^\theta_{u_3})|X^\theta_{t_i}=x]du_3du_2du_1.$$

We now study $A^2h_{i,n}(x)$. We recall it is

$$A^2h_{i,n}(x)=\bar{A}_c^2h_{i,n}(x)+\bar{A}_cA_dh_{i,n}(x)+A_d\bar{A}_ch_{i,n}(x)+A_d^2h_{i,n}(x).$$

We observe that we can write $\bar{A}_c^2h_{i,n}(x)$ as $\sum_{j=1}^4h_j(x,\theta)h^{(j)}_{i,n}(x)$, where, for each $j\in\{1,2,3,4\}$, h_j is a function of a, $\bar{b}$ and their derivatives up to second order: $h_1=\frac{1}{2}a^2\bar{b}''+\bar{b}\bar{b}'$, $h_2=\frac{1}{2}a^2(a')^2+\frac{1}{2}a^3a''+a^2\bar{b}'+aa'\bar{b}+\bar{b}^2$, $h_3=a^3a'+a^2\bar{b}$ and $h_4=\frac{1}{4}a^4$. Moreover, recalling that by the definition of φ we have $\varphi_{\Delta_{n,i}^\beta}(0)=1$ and $\varphi^{(k)}_{\Delta_{n,i}^\beta}(0)=0$ $\forall k\geq 1$, it follows $h'_{i,n}(x)=2(x-m(\mu,\sigma,x))$, $h''_{i,n}(x)=2$ and $h^{(l)}_{i,n}=0$ $\forall l\geq 3$. We obtain

$$\bar{A}_c^2h_{i,n}(x)=2h_1(x,\theta)(x-m(\mu,\sigma,x))+2h_2(x,\theta)=R(\theta,\Delta_{n,i},x)+2h_2(x,\theta), \quad (145)$$

where the last equality is a consequence of the development (31) of m.
Concerning $\bar{A}_cA_dh_{i,n}(x)$, it is $\frac{1}{2}a^2(x,\sigma)(A_dh_{i,n}(x))''+\bar{b}(x,\mu)(A_dh_{i,n}(x))'$. We start considering

$$(A_dh_{i,n}(x))'=\int_{\mathbb{R}}[h'_{i,n}(x+z\gamma(x))(1+z\gamma'(x))-h'_{i,n}(x)]F(z)dz.$$

We now observe that, $\forall k\geq 1$ and $\forall y\in\mathbb{R}$, $|y-m(\mu,\sigma,x)|^k|\varphi_{\Delta_{n,i}^\beta}(y-x)|\leq(|y-x|^k+|x-m(\mu,\sigma,x)|^k)|\varphi_{\Delta_{n,i}^\beta}(y-x)|$. We have that $|y-x|^k|\varphi_{\Delta_{n,i}^\beta}(y-x)|\leq c\Delta_{n,i}^{k\beta}|\varphi_{\Delta_{n,i}^\beta}(y-x)|$ as a consequence of the definition of φ

while, using the development (31) of m , it follows $|x-m(\mu,\sigma,x)|^k|\varphi_{\Delta^\beta_{n,i}}(y-x)| \le R(\theta,\Delta^k_{n,i},x)$. Putting the pieces together it is

$$|y-m(\mu,\sigma,x)|^k|\varphi_{\Delta^\beta_{n,i}}(y-x)| \le R(\theta,\Delta^{\beta k}_{n,i},x)+R(\theta,\Delta^k_{n,i},x)=R(\theta,\Delta^{\beta k}_{n,i},x); \tag{146}$$

it is easy to remark that the same reasoning holds with the derivatives of φ instead of φ. We underline that from the estimation (146) here above and the computation of the derivatives of $h_{i,n}$ we get that, $\forall l \ge 0$, each term of the l-derivative of $h_{i,n}$ has the same size. Indeed, each time we derive φ an extra $\Delta^{-\beta}_{n,i}$ turns out but we can recover it on the basis of (146). In particular, $\forall l \ge 0$ it follows

$$\left\|h^{(l)}_{i,n}\right\|_\infty \le R(\theta,\Delta^{\beta(2-l)}_{n,i},x). \tag{147}$$

Therefore we obtain

$$|(A_d h_{i,n}(x))'| \le c\Delta^\beta_{n,i}(2\lambda + c\left\|\gamma'\right\|_\infty). \tag{148}$$

Concerning the second derivative of $A_d h_{i,n}(x)$, it is

$$(A_d h_{i,n}(x))'' = \int_{\mathbb{R}} [h''_{i,n}(x+z\gamma(x))(1+z\gamma'(x))^2 + h'_{i,n}(x+z\gamma(x))z\gamma''(x) - h''_{i,n}(x)]F(z)dz =$$

$$= \int_{\mathbb{R}} [h''_{i,n}(x+z\gamma(x)) - h''_{i,n}(x) + 2z\gamma'(x)h''_{i,n}(x+z\gamma(x)) + z^2(\gamma'(x))^2 h''_{i,n}(x+z\gamma(x)) + h'_{i,n}(x+z\gamma(x))z\gamma''(x)]F(z)dz = \sum_{j=1}^5 I_j.$$

On I_3 and I_4 we act like we did on the integral in (141), observing that in the computation of $h''_{i,n}$ we always have φ or its derivatives which make the integrals different from 0 only for $|z| \le c\Delta^\beta_{n,i}$. On I_5 we use (147) for $l=1$, getting $|I_5| \le R(\theta,\Delta^\beta_{n,i},x)$. We observe that, by the computation of $h''_{i,n}(x)$ we obtain $I_2 = -2\lambda$.
To conclude the study of $\bar{A}_c A_d h_{i,n}(x)$ we have to deal with I_1:

$$\int_{\mathbb{R}} h''_{i,n}(x+z\gamma(x))F(z)dz = \int_{\mathbb{R}} 2\varphi_{\Delta^\beta_{n,i}}(z\gamma(x))F(z)dz + 4\Delta^{-\beta}_{n,i}(x-m(\mu,\sigma,x))\int_{\mathbb{R}} \varphi'_{\Delta^\beta_{n,i}}(z\gamma(x))F(z)dz + \tag{149}$$
$$+4\Delta^{-\beta}_{n,i}\int_{\mathbb{R}} z\gamma(x)\varphi'_{\Delta^\beta_{n,i}}(z\gamma(x))F(z)dz + \Delta^{-2\beta}_{n,i}\int_{\mathbb{R}}((x-m(\mu,\sigma,x))+z\gamma(x))^2\varphi''_{\Delta^\beta_{n,i}}(z\gamma(x))F(z)dz.$$

Applying the change of variable $u := \frac{z\gamma(x)}{\Delta^\beta_{n,i}}$ and recalling that from the development of m it follows $|x-m(\mu,\sigma,x)|^k \le R(\theta,\Delta^k_{n,i},x)$ for each $k \ge 1$, we obtain

$$\int_{\mathbb{R}} h''_{i,n}(x+z\gamma(x))F(z)dz = \frac{\Delta^\beta_{n,i}}{\gamma(x)}\int_{\mathbb{R}}(2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta^\beta_{n,i}}{\gamma(x)})du + R(\theta,\Delta^{(1-\beta)\wedge(2-2\beta)\wedge(1-2\beta)}_{n,i},x).$$

It is worth noting that the magnitude of the first term in the left hand side of the equation here above depends on the density F.
Remarking also that $\varphi(u)=0$ for each u such that $|u| \ge 2$, it follows

$$(A_d h_{i,n}(x))'' = \frac{\Delta^\beta_{n,i}}{\gamma(x)}\int_{u:|u|\le 2}(2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta^\beta_{n,i}}{\gamma(x)})du - 2\lambda + R(\theta,\Delta^{(1-2\beta)\wedge\beta}_{n,i},x). \tag{150}$$

From the definition of $\bar{A}_c A_d h_{i,n}(x)$, (148) and (150) we get

$$\bar{A}_c A_d h_{i,n}(x) = \frac{\Delta^\beta_{n,i}a^2(x,\sigma)}{2\gamma(x)}\int_{\mathbb{R}}(2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta^\beta_{n,i}}{\gamma(x)})du - a^2(x,\sigma)\lambda + R(\theta,\Delta^{(1-2\beta)\wedge\beta}_{n,i},x). \tag{151}$$

Now we deal with

$$A_d\bar{A}_c h_{i,n}(x) = \int_{\mathbb{R}}[\bar{A}_c h_{i,n}(x+z\gamma(x)) - \bar{A}_c h_{i,n}(x)]F(z)dz. \tag{152}$$

From (143) it follows

$$\int_{\mathbb{R}} \bar{A}_c h_{i,n}(x)F(z)dz = \lambda a^2(x,\sigma) + 2\lambda(x-m(\mu,\sigma,x))\bar{b}(x,\mu) = \lambda a^2(x,\sigma) + R(\theta,\Delta_{n,i},x), \tag{153}$$

where we have also used the development (31) of m. Moreover,

$$\int_{\mathbb{R}} \bar{A}_c h_{i,n}(x+z\gamma(x))F(z)dz = \int_{\mathbb{R}}[\frac{1}{2}a^2(x+z\gamma(x),\sigma)h''_{i,n}(x+z\gamma(x)) + \bar{b}(x+z\gamma(x),\mu)h'_{i,n}(x+z\gamma(x))]F(z)dz. \tag{154}$$

Acting on the first term of the right hand side of the equation here above exactly like we did in (149) we get it is equal to

$$\frac{\Delta_{n,i}^{\beta}}{2\gamma(x)}\int_{u:|u|\leq 2} a^2(x+u\Delta_{n,i}^{\beta},\sigma)(2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du + R(\theta,\Delta_{n,i}^{1-2\beta},x). \tag{155}$$

We upper bound the second term of the right hand side of (154), instead, using (147) for $l=1$. It yields

$$|\int_{\mathbb{R}} \bar{b}(x+z\gamma(x),\mu)h'_{i,n}(x+z\gamma(x))F(z)dz| \leq R(\theta,\Delta_{n,i}^{\beta},x), \tag{156}$$

where we have also used the polynomial growth of b. Replacing in (152) the equations (153) - (156) we obtain

$$A_d\bar{A}_c h_{i,n}(x) = \frac{\Delta_{n,i}^{\beta}}{2\gamma(x)}\int_{u:|u|\leq 2} a^2(x+u\Delta_{n,i}^{\beta},\sigma)(2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du - \lambda a^2(x,\sigma) + R(\theta,\Delta_{n,i}^{(1-2\beta)\wedge\beta},x). \tag{157}$$

We have to study $A_dA_dh_{i,n}(x)$. It is

$$|A_d^2h_{i,n}(x)| = |\int_{\mathbb{R}}[A_dh_{i,n}(x+z\gamma(x))-A_dh_{i,n}(x)]F(z)dz| \leq \int_{\mathbb{R}} \|(A_dh_{i,n})'\|_{\infty}|z\gamma(x)|F(z)dz \leq R(\theta,\Delta_{n,i}^{\beta},x), \tag{158}$$

where we have used the definition $A_dh'_{i,n}$ and (147), remarking that the estimation (148) holds also in no matter which $y\in\mathbb{R}$.
From (145), (151), (157) and (158) it follows

$$\frac{1}{2}\Delta_{n,i}^2A^2h_{i,n}(x) = \Delta_{n,i}^2h_2(x,\theta) - \lambda\Delta_{n,i}^2a^2(x,\sigma) + \tag{159}$$

$$+\frac{\Delta_{n,i}^{2+\beta}}{4\gamma(x)}\int_{u:|u|\leq 2}[a^2(x,\sigma)+a^2(x+u\Delta_{n,i}^{\beta},\sigma)](2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du + R(\theta,\Delta_{n,i}^{(3-2\beta)\wedge(2+\beta)},x).$$

To complete the study of the numerator of m_2 we need to estimate $\frac{1}{6}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}\int_{t_i}^{u_2}\mathbb{E}[A^3h_{i,n}(X_{u_3}^{\theta})|X_{t_i}^{\theta}=x]du_3du_2du_1$. We introduce the following norm on $\mathcal{C}^p$ functions, with $p\geq 0$, $c>0$:

$$\|f\|_{\infty,c,p} := \sum_{k=0}^{p}\sup_{y\in\mathbb{R}}|\frac{f^{(k)}(y)}{(1+|y|)^c}|.$$

We observe it is

$$\int_{\mathbb{R}}|f(y+z\gamma(y))|F(z)dz = \int_{\mathbb{R}}\frac{|f(y+z\gamma(y))|(1+|y+z\gamma(y)|)^{\tilde{c}}}{(1+|y+z\gamma(y)|)^{\tilde{c}}}F(z)dz \leq \|f\|_{\infty,\tilde{c},0}\int_{\mathbb{R}}(1+|y+z\gamma(y)|)^{\tilde{c}}F(z)dz.$$

We can therefore evaluate the norm of A_df, getting

$$\|A_df\|_{\infty,\tilde{c},0} = \left\|\frac{A_df}{(1+|y|)^{\tilde{c}}}\right\|_{\infty} \leq c\|f\|_{\infty,\tilde{c},0}\,(\frac{\int_{\mathbb{R}}(1+|y|^{\tilde{c}}+(1+|y|)^{\tilde{c}}|z|^{\tilde{c}})F(z)dz}{(1+|y|)^{\tilde{c}}}+\lambda) \leq c\|f\|_{\infty,\tilde{c},0}\,.$$

By similar computations on the derivatives of A_df we obtain, $\forall p\geq 0$, with $p\leq 4$, $\|A_df\|_{\infty,\tilde{c},p}\leq c\|f\|_{\infty,\tilde{c},p}$ (similar calculations are in Theorem 2.3 of [30]).
In order to find an upper bound for the norm of $\bar{A}_cf$ we observe that, from the polynomial growth of both the coefficients a and b, it follows

$$|\frac{\bar{A}_cf}{(1+|y|)^{\tilde{c}+2}}| = |\frac{\frac{1}{2}a^2(y,\sigma)f''(y)}{(1+|y|)^{\tilde{c}+2}}+\frac{b(y,\mu)f'(y)}{(1+|y|)^{\tilde{c}+2}}| \leq \frac{c}{(1+|y|)^{\tilde{c}}}|f''(y)|+\frac{c}{(1+|y|)^{\tilde{c}}}|f'(y)|.$$

Hence, we deduce that $\left\|\bar{A}_cf\right\|_{\infty,\tilde{c}+2,0}\leq c\|f\|_{\infty,\tilde{c},2}$. Acting again with similar computation on the following derivatives, as done detailing in the proof of Theorem 2.3 in [30], we get $\left\|\bar{A}_cf\right\|_{\infty,\tilde{c}+2,p}\leq c\|f\|_{\infty,\tilde{c},p+2}$.
We want to use the estimations on $\|A_df\|_{\infty,\tilde{c},p}$ and $\left\|\bar{A}_cf\right\|_{\infty,\tilde{c},p}$ and equation (147) to evaluate each term of
$\mathbb{E}[A^3h_{i_n}(X_{u_3}^{\theta})|X_{t_i}^{\theta}=x]$ but $\bar{A}_c^3h_{i,n}$. We observe it is, for $\tilde{c}\geq 4$,

$$\mathbb{E}_i[A_d^3h_{i,n}(X_{u_3}^{\theta})] \leq \left\|A_d^3h_{i,n}\right\|_{\infty,\tilde{c},0}\mathbb{E}_i[(1+|X_{u_3}^{\theta}|)^{\tilde{c}}] \leq c\|h_{i,n}\|_{\infty,\tilde{c},0}\,R(\theta,1,x) \leq R(\theta,\Delta_{n,i}^{2\beta},x);$$

$$\mathbb{E}_i[A_dA_d\bar{A}_ch_{i,n}(X_{u_3}^{\theta})] \leq \left\|A_dA_d\bar{A}_ch_{i,n}\right\|_{\infty,\tilde{c},0}\mathbb{E}_i[(1+|X_{u_3}^{\theta}|)^{\tilde{c}}] \leq c\|h_{i,n}\|_{\infty,\tilde{c}-2,2}\,R(\theta,1,x) \leq R(\theta,1,x);$$

remarking that the same estimation holds for $A_d\bar{A}_cA_dh_{i_n}$ and $\bar{A}_cA_dA_dh_{i_n}$. Moreover we have

$$\mathbb{E}_i[\bar{A}_c\bar{A}_cA_dh_{i,n}(X^\theta_{u_3})] \leq \left\|\bar{A}_c\bar{A}_cA_dh_{i,n}\right\|_{\infty,\tilde{c},0}\mathbb{E}_i[(1+|X^\theta_{u_3}|)^{\tilde{c}}] \leq c\left\|h_{i,n}\right\|_{\infty,\tilde{c}-4,4}R(\theta,1,x) \leq R(\theta,\Delta_{n,i}^{-2\beta},x)$$

and we can upper bound $\bar{A}_cA_d\bar{A}_ch_{i,n}$ and $A_d\bar{A}_c\bar{A}_ch_{i,n}$ with the same quantity.
We are now left to study $\mathbb{E}_i[\bar{A}_c^3h_{i,n}(X^\theta_{u_3})]$. As we have already done considering $\bar{A}_c^2h_{i,n}(x)$, we see $\bar{A}_c^3h_{i,n}$ as linear combination of the derivatives of $h_{i,n}$: $\bar{A}_c^3h_{i,n}(y) := \sum_{j=1}^6 h_j(y,\theta)h_{i,n}^{(j)}(y)$ where for each $j \in \{1,...,6\}$ h_j is a function of a, b and their derivatives up to fourth order.
Using a conditional version of Proposition 11 we get $\forall\epsilon > 0$, $\forall j \geq 3$ $\mathbb{E}_i[|h_j(X^\theta_{u_3},\theta)h_{i,n}^{(j)}(X^\theta_{u_3})|] \leq R(\theta,\Delta_{n,i}^{1+(2-j)\beta-\epsilon},x)$.
Concerning the first two terms of the sum, we have

$$|\mathbb{E}_i[h_1(X^\theta_{u_3},\theta)h'_{i,n}(X^\theta_{u_3})+h_2(X^\theta_{u_3},\theta)h''_{i,n}(X^\theta_{u_3})]| \leq \left\|h'_{i,n}\right\|_\infty\mathbb{E}_i[|h_1(X^\theta_{u_3},\theta)|]+\left\|h''_{i,n}\right\|_\infty\mathbb{E}_i[|h_2(X^\theta_{u_3},\theta)|] \leq R(\theta,1,x),$$

which follows from (147) and from the polynomial growth of a, b and their derivatives, which constitute the functions h_1 and h_2. Putting all the pieces together we get

$$\frac{1}{6}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}\int_{t_i}^{u_2}\mathbb{E}[A^3h_{i,n}(X^\theta_{u_3})|X^\theta_{t_i}=x]du_3du_2du_1 \leq R(\theta,\Delta_{n,i}^{(3-2\beta)\wedge(4-4\beta-\epsilon)},x) = R(\theta,\Delta_{n,i}^{3-2\beta},x).$$

From (144), (159) and the equation here above it follows $\mathbb{E}_i[h_{i,n}(X^\theta_{t_{i+1}})] =$

$$= (x-m(\mu,\sigma,x))^2+\Delta_{n,i}a^2(x,\sigma)+\frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)}\int_{\mathbb{R}}v^2\varphi(v)F(\frac{v\Delta_{n,i}^\beta}{\gamma(x)})dv+2\Delta_{n,i}(x-m(\mu,\sigma,x))\bar{b}(x,\mu)+\Delta_{n,i}^2h_2(x,\theta)+ \tag{160}$$

$$-\lambda\Delta_{n,i}^2a^2(x,\sigma)+\frac{\Delta_{n,i}^{2+\beta}}{4\gamma(x)}\int_{u:|u|\leq 2}[a^2(x,\sigma)+a^2(x+u\Delta_{n,i}^\beta,\sigma)](2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)})du+R(\theta,\Delta_{n,i}^{(3-2\beta)\wedge(2+\beta)},x).$$

The expansion (137) follows from (160) and (16), with the smoothness of $z \mapsto a^2(z,\sigma)$, and $\int_{\mathbb{R}}F(z)dz < \infty$.

□

Proof of (138). Concerning the denominator of m_2, we still use Dynkin formula up to third order, this time on $f_{i,n}(y) := \varphi_{\Delta_{n,i}^\beta}(y-x)$. We observe that, by the building, $f_{i,n}(x) = 1$ and $f_{i,n}^{(k)}(x) = 0$ for each $k \geq 1$. Hence, $\bar{A}_cf_{i,n}(x) = 0$ and

$$A_df_{i,n}(x) = \int_{\mathbb{R}}[f_{i,n}(x+\gamma(x)z)-1]F(z)dz = \int_{z:|z|\leq\frac{2\Delta_{n,i}^\beta}{|\gamma(x)|}}\varphi_{\Delta_{n,i}^\beta}(z\gamma(x))F(z)dz-\lambda.$$

As we have already done we can see the first term here above, after the change of variable $u := \frac{z\gamma(x)}{\Delta_{n,i}^\beta}$, as $\frac{\Delta_{n,i}^\beta}{\gamma(x)}\int_{u:|u|\leq 2}\varphi(u)F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)})du$, which order depends on the density F.
Concerning the study of $A^2f_{i,n}(x)$, we first of all remark that $\bar{A}_c^2f_{i,n}(x) = 0$.
Moreover, we observe it is $f_{i,n}^{(k)}(y) = \Delta_{n,i}^{-\beta k}\varphi_{\Delta_{n,i}^\beta}^{(k)}(y-x)$ and so by the boundedness of φ and its derivatives we get that, for each $k \geq 1$,

$$\left\|f_{i,n}^{(k)}\right\|_\infty \leq R(\theta,\Delta_{n,i}^{-\beta k},x). \tag{161}$$

We therefore have, $\forall y \in \mathbb{R}$,

$$|(A_df_{i,n}(y))'| = |\int_{\mathbb{R}}f'_{i,n}(y+z\gamma(y))(1+z\gamma'(y))F(z)dz-\lambda f'_{i,n}(y)| \leq R(\theta,\Delta_{n,i}^{-\beta},y)$$

and, in the same way, $|(A_df_{i,n}(y))''| \leq R(\theta,\Delta_{n,i}^{-2\beta},y)$. It follows

$$|\bar{A}_cA_df_{i,n}(x)| = |\frac{1}{2}a^2(x,\sigma)(A_df_{i,n}(x))''+b(x,\mu)(A_df_{i,n}(x))'| \leq R(\theta,\Delta_{n,i}^{-2\beta},x);$$

$$|A_d\bar{A}_cf_{i,n}(x)| = |\int_{\mathbb{R}}\bar{A}_cf_{i,n}(x+z\gamma(x))F(z)dz| \leq R(\theta,\Delta_{n,i}^{-2\beta},x)$$

and, using also finite-increments theorem, $|A_dA_df_{i,n}(x)| \leq R(\theta,\Delta_{n,i}^{-\beta},x)$. Putting pieces together we have $|\frac{1}{2}\Delta_{n,i}^2A^2f_{i,n}(x)| \leq R(\theta,\Delta_{n,i}^{2-2\beta},x)$.
Considering the integral rest of the Dynkin formula , we act like we did in the study of the numerator,

passing through the use of the norm $\|.\|_{\infty,c,p}$ and the estimation of the derivatives of $f_{i,n}$ gathered in (161). It yields

$$\mathbb{E}_i[A_d^3 f_{i_n}(X_{u_3}^\theta)] \leq R(\theta,1,x), \qquad \mathbb{E}_i[A_d A_d \bar{A}_c f_{i_n}(X_{u_3}^\theta)] \leq R(\theta,\Delta_{n,i}^{-2\beta},x),$$

$$\mathbb{E}_i[A_d \bar{A}_c \bar{A}_c f_{i_n}(X_{u_3}^\theta)] \leq R(\theta,\Delta_{n,i}^{-4\beta},x), \qquad \mathbb{E}_i[\bar{A}_c^3 f_{i_n}(X_{u_3}^\theta)] \leq R(\theta,\Delta_{n,i}^{1-6\beta-\epsilon},x),$$

having also used Proposition 11 to get the last one estimation here above. It turns out the denominator is

$$1-\Delta_{n,i}\lambda+\frac{\Delta_{n,i}^{1+\beta}}{\gamma(x)}\int_{u:|u|\leq 2}\varphi(u)F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du+R(\theta,\Delta_{n,i}^{2-2\beta},x); \tag{162}$$

since we can always find an $\epsilon>0$ for which $4-6\beta-\epsilon>2-2\beta$ and we have $3-4\beta>2-2\beta$. This concludes the proof of the expansion (138). □

To conclude the proof of the Proposition 4 we are left to show that the derivatives with respect to both the parameters of the rest terms in the expansions (137)–(138) are still rest functions and their order remains the same.
We observe that up to the development of second order in Dynkin formula, the rest functions R are totally explicit in our computation and so it is possible to calculate its derivatives with respect to both μ and σ. As we have already seen during the proof, we can use the estimations on $\partial_\vartheta m(\mu,\sigma,x)$ for $\vartheta=\mu$ and $\vartheta=\sigma$ gathered in Proposition 8 and the fact that $(x-m(\mu,\sigma,x))$ is a $R(\theta,\Delta_{n,i},x)$ function to get that size of $h_{i,n}$ and of the rest functions does not change after having derived with respect to the parameters. Concerning the integral rest coming from the third order of the Dynkin formula, we have that

$$\partial_\vartheta \mathbb{E}_i[A^3 h_{i,n}(X_{u_3}^\theta)] = \mathbb{E}_i[\partial_\vartheta A^3 h_{i,n}(X_{u_3}^\theta)] + \mathbb{E}_i[\partial_X A^3 h_{i,n}(X_{u_3}^\theta)\partial_\vartheta X].$$

On the first term of the right hand side here above we can act exactly like we did on $\mathbb{E}_i[A^3 h_{i,n}(X_{u_3}^\theta)]$ getting a rest function whose order does not change, while from the computation of $\partial_X A^3 h_{i,n}(X_{u_3}^\theta)$ an extra $\Delta_{n,i}^{-\beta}$ appears but, since from Lemma 14 the norm 1 of $\partial_\vartheta X$ is $R(\theta,\Delta_{n,i},x)$ for $\vartheta=\mu$ and $R(\theta,\Delta_{n,i}^{\frac{1}{2}},x)$ for $\vartheta=\sigma$, it is enough to use previously Holder inequality and observe that both $\frac{1}{2}-\beta$ and $1-\beta$ are positive to get that the second term here above is negligible compared to the first.

A.3.3 Proof of Proposition 3

Proposition 3 is a particular case in which Proposition 4 holds. The proof relies on the fact that the intensity F is supposed to be $\mathcal{C}^1$ and so we can move from $F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})$ to $F(0)$ through finite-increments theorem.

Proof. From (15) we get the proposition proved remarking that

$$\Delta_{n,i}^{1+2\beta}\int_{\mathbb{R}}u^2\varphi(u)F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du=\Delta_{n,i}^{1+2\beta}\int_{\mathbb{R}}u^2\varphi(u)F(0)du+\Delta_{n,i}^{1+2\beta}\int_{\mathbb{R}}u^2\varphi(u)[F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})-F(0)]du$$

and, from finite-increments theorem, the last term is in absolute value upper bounded by
$c\Delta_{n,i}^{1+3\beta}\frac{1}{|\gamma(x)|}\int_{\mathbb{R}}|u|^3|\varphi(u)||F'(\tilde{u})|du=R(\theta,\Delta_{n,i}^{1+3\beta},x)$, where $\tilde{u}\in(0,\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})$.
Moreover, by the smoothness on F we have required, it follows that terms in (15) whose size depends on the density F are now upper bounded by a $R(\theta,\Delta_{n,i}^{2+\beta},x)$ function. It yields (14). □

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] Aït-Sahalia, Y. and Yu, J. (2006). Saddlepoint approximations for continuous-time Markov processes, Journal of Econometrics, 134, 507551

[2] Amorino C, and Gloter A. (2019) Contrast function estimation for the drift parameter of ergodic jump diffusion process. Scand J Statist. 2019;168. https://doi.org/10.1111/sjos.12406.

[3] Amorino, C., Gloter, A. (2020). Unbiased truncated quadratic variation for volatility estimation in jump diffusion processes. Stochastic Processes and their Applications.

[4] Applebaum, David. Lévy processes and stochastic calculus. Cambridge university press, 2009.

[5] Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. J. R. Stat. Soc., Ser. B, Stat. Methodol., 63, 167-241.

[6] Bates, D.S. (1996). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark. The Review of Financial Studies, 9(1), 69-107.

[7] Bichteler, K., Gravereaux, J. B. and Jacod, J. Malliavin calculus for processes with jumps, Gordon and Breach Science Publishers, New York, 1987. MR MR1008471 (90h: 60056).

[8] Comte F, Genon-Catalot V. Nonparametric estimation for pure jump Lévy processes based on high frequency data. Stochastic Processes and their Applications. 2009 Dec 1;119(12):4088-123.

[9] Ditlevsen, S. and Greenwood, P. (2013). The MorrisLecar neuron model embeds a leaky integrate-and-fire model. Journal of Mathematical Biology 67 239-259.

[10] Doukhan, P. (2012). Mixing: properties and examples (Vol. 85). Springer Science and Business Media.

[11] Eraker, B., Johannes, M. and Polson N. (2003). The Impact of Jumps in Volatility and Returns. J. Finance, 58(3), 1269-1300.

[12] Figueroa-López, J. E. (2011). Sieve-based confidence intervals and bands for Lévy densities. Bernoulli, 17(2), 643-670.

[13] Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. Statistics: A Journal of Theoretical and Applied Statistics, 20(4), 547-557.

[14] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. Annales de linstitut Henri Poincar (B) Probabilits et Statistiques, 29, 119-151.

[15] Gloter, A., Loukianova, D. and Mai, H. (2018). Jump filtering and efficient drift estimation for Lvy-driven SDEs. The Annals of Statistics, 46(4), 1445-1480.

[16] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. In Annales de l'IHP Probabilités et statistiques (Vol. 38, No. 5, pp. 711-737).

[17] Gugushvili, S. (2012). Nonparametric inference for discretely sampled Lévy processes. In Annales de l'IHP Probabilits et statistiques (Vol. 48, No. 1, pp. 282-307).

[18] Hall, P. and Heyde, C. (1980). Martingale Limit Theory and its Applications,Academic Press, New York.

[19] Iacus, S.M. and Yoshida, N., 2017. Simulation and inference for stochastic processes with YUIMA. Springer Series in Statistics, Springer NY.

[20] Inatsugu, H. and Yoshida, N., 2018. Global jump filters and quasi likelihood analysis for volatility. arXiv preprint arXiv:1806.10706.

[21] Jacod, J., and Protter, P. (2011). Discretization of processes (Vol. 67). Springer Science and Business Media.

[22] Jakobsen, N. and Sørensen, M. (2017). Estimating functions for jump-diffusions, Preprint.

[23] Jongbloed G, Van Der Meulen FH, Van Der Vaart AW. Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. Bernoulli. 2005;11(5):759-91.

[24] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. Scandinavian Journal of Statistics, 24(2), 211-229.

[25] Kou, S.G. (2002). A Jump-Diffusion Model for Option Pricing. Management Science, 48, 1086-1101.

[26] Li, C. and Chen, D. (2016). Estimating jump-diffusions using closed form likelihood expansions. Journal of Econometrics, 195, 5171

[27] Masuda, H. (2007). Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. Stochastic processes and their applications, 117(1), 35-56.

[28] Masuda, H. (2013). Convergence of gaussian quasi-likelihood random fields for ergodic lvy driven sde observed at high frequency. Annals of stat., 41(3), 1593-1641.

[29] Merton, R.C. (1976). Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics, 3, 125-144

[30] Mies, F. (2018). State-dependent jump activity estimation for Markovian semimartingales. arXiv preprint arXiv:1811.06351.

[31] Nikolskii, S.M. (1977). Approximation of functions of several variables and imbedding theorems (Russian). Sec. ed., Moskva, Nauka 1977 English translation of the first ed., Berlin 1975

[32] Protter, P. E. (2004). Stochastic integration and differential equations. 2nd ed. Applications of Mathematics 21. Berlin: Springer.

[33] Shimizu, Y. (2007). Asymptotic Inference for Stochastic Differential Equations with Jumps from Discrete Observations and Some Practical Approaches (Doctoral dissertation, University of Tokyo).

[34] Shimizu, Y. (2006). M-Estimation for Discretely Observed Ergodic Diffusion Processes with Infinitely many Jumps. Statistical Inference for Stochastic Processes, 9, 179-225.

[35] Shimizu, Y. (2009). A practical inference for discretely observed jump-diffusions from finite samples. Journal of the Japan Statistical Society, 38(3), 391-413.

[36] Shimizu, Y., and Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. Statistical Inference for Stochastic Processes, 9(3), 227-277.

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