



HAL
open science

Estimation of parameters and local times in a discretely observed threshold diffusion model

Sara Mazzonetto, Paolo Pigato

► **To cite this version:**

Sara Mazzonetto, Paolo Pigato. Estimation of parameters and local times in a discretely observed threshold diffusion model. 2024. hal-04500135

HAL Id: hal-04500135

<https://hal.science/hal-04500135>

Preprint submitted on 11 Mar 2024

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Estimation of parameters and local times in a discretely observed threshold diffusion model

Sara Mazzonetto* and Paolo Pigato†

March 11, 2024

Abstract

We consider a simple mean reverting diffusion process, with piecewise constant drift and diffusion coefficients, discontinuous at a fixed threshold. We discuss estimation of drift and diffusion parameters from discrete observations of the process, with a generalized moment estimator and a maximum likelihood estimator. We develop the asymptotic theory of the estimators when the time horizon of the observations goes to infinity, considering both cases of a fixed time lag (low frequency) and a vanishing time lag (high frequency) between consecutive observations. In the setting of low frequency observations and infinite time horizon we also study the convergence of three local time estimators, that are already known to converge to the local time in the setting of high frequency observations and fixed time horizon. We find that these estimators can behave differently, depending on the assumptions on the time lag between observations.

Keywords: Threshold diffusion, maximum likelihood, generalised moment estimator, local time, discrete observations.

AMS 2020: primary: 62F12; secondary: 62M05; 60F05; 60J55.

1 Introduction

We consider the diffusion process solution to the following stochastic differential equation (SDE)

$$X_t = X_0 + \int_0^t \sigma(X_s) dW_s + \int_0^t b(X_s) ds, \quad t \geq 0, \quad (1.1)$$

with piecewise constant volatility and drift coefficient, possibly discontinuous at $r \in \mathbb{R}$,

$$\sigma(x) = \begin{cases} \sigma_+ > 0 & \text{if } x \geq r, \\ \sigma_- > 0 & \text{if } x < r, \end{cases} \quad \text{and} \quad b(x) = \begin{cases} b_+ \in \mathbb{R} & \text{if } x \geq r, \\ b_- \in \mathbb{R} & \text{if } x < r. \end{cases}$$

We assume the initial condition to be deterministic. Separately on (r, ∞) and $(-\infty, r)$, the process follows the drifted Brownian motion dynamics with different parameters. We also assume that $b_- > 0$ and $b_+ < 0$, so that the process is mean reverting and ergodic.

*Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France. E-mail: sara.mazzonetto@univ-lorraine.fr

†Department of Economics and Finance, University of Rome Tor Vergata, Via Columbia 2, 00133 Roma, Italy. E-mail: paolo.pigato@uniroma2.it

We discuss the estimation of parameters from discrete observations of X , when the threshold r is known. We consider two types of drift estimator: a generalized moment estimator (GME) and a maximum discretized likelihood estimator (dMLE). We prove that they are equivalent in a suitable sense, so that the asymptotic theory we develop applies to both.

We construct the GME, based on discrete observations, using the theoretical long time behavior of conditional moments and local times of X . Maximum likelihood estimation (MLE) of this process has been considered in Lejay and Pigato [2020], assuming that X is continuously monitored and the observation span goes to infinity. Since we assume here to observe the process at discrete times, we study the corresponding dMLE, with observation span going to infinity.

We consider two types of asymptotic setting: (a) fixed time lag h (low frequency observations) and number of observations N going to infinity, and (b) shrinking time lag h_N (high frequency observations) and simultaneously time horizon $T_N = Nh_N$ going to infinity. In all cases we prove consistency and asymptotic normality of GME and dMLE estimators. We also show that the dMLE based on N observations converges in high frequency to the MLE based on continuous observations, with speed $N^{1/4}$, and an analogous result for the GME. Moreover, in setting (a), we propose a GME for the diffusion coefficient, for which we prove again consistency and asymptotic normality.

Finally, we consider three different estimators for the local time at the discontinuity level r , that are known to converge to the local time with speed $N^{1/4}$ in the high frequency limit of the observations. We show that when the observation lag is fixed and the time horizon goes to infinity, the asymptotic behavior is the same as for the local time only for one of the estimators, while the other two have different limit behaviors, that depend on the observation lag.

Related work. The solution to (1.1) with piecewise affine instead of piecewise constant drift is the so-called threshold Ornstein-Uhlenbeck (OU) process, for which parameters inference (including the threshold r) has been discussed e.g. in Kutoyants [2012], Dieker and Gao [2013], Su and Chan [2015, 2017], Yu et al. [2020], Hu and Xi [2022], Mazzonetto and Pigato [2024]. Estimation of a piecewise constant drift as in (1.1), despite the process having several applications (see below), has been considered less, see e.g. Mota and Esquível [2014], Lejay and Pigato [2020].

In the low frequency observations setting, the main tool we use in this paper is the ergodic theorem in the form of [Meyn and Tweedie, 2009, Theorem 17.0.1], in the multidimensional framework as in [Brooks et al., 2011, Section 1.8], in the spirit of Hu and Xi [2022]. In the high frequency observations setting, we obtain that the asymptotic behavior is the same as for the continuous observations case, under the condition $\lim_{N \rightarrow \infty} h_N T_N = 0$, analogous to the condition given for diffusions with no threshold in [Kessler, 1997, Ben Alaya and Kebaier, 2013, Amorino and Gloter, 2020] and with threshold in Mazzonetto and Pigato [2024].

As already mentioned, several local time estimators for Brownian motion or skew Brownian motion, with fixed time horizon and N observations, have been considered in the literature (see e.g. Portenko [1994], Borodin [1986], Lejay and Pigato [2018, 2020], Lejay et al. [2019]) and they are known to converge to the local time with speed $N^{1/4}$ in the high-frequency limit of the observations [Jacod, 1998, Mazzonetto, 2019]. Local time approximation is an important topic in statistics of processes, for instance because the amount of time spent at a certain level is related to the accuracy of the estimation of the process at that level [Florens-Zmirou, 1993]. Recently in Christensen and Strauch [2023], Christensen et al. [2023] the local time of a scalar diffusion model have been used in the exploration phase of studies of the exploration vs exploitation tradeoff in a reinforcement learning setting. The authors point out the difficulties related to local time estimation and propose strategies to avoid it.

The process in (1.1) and variations of it have been widely used in financial modelling, from the point of view of time series (see e.g. Ang and Timmermann [2012], Mota and Esquível

[2014], Lejay and Pigato [2019]), options pricing and implied volatility (see e.g. Lipton and Sepp [2011-10], Gairat and Shcherbakov [2016], Dong and Wong [2017], Lipton [2018], Pigato [2019], Buckner et al. [2024]), interest rates (see e.g. Pai and Pedersen [1999], Decamps et al. [2006], Su and Chan [2015, 2017]) and others. The discrete time analog of threshold diffusions are Threshold autoregressive (TAR) and in particular self-exciting (SETAR) models [Tong, 2011, Chen et al., 2011]. They have been widely used in financial modelling, recently also in combination with reinforcement learning in Giorgi et al. [2023]. Threshold processes have also a wide range of applications outside of financial modelling; for example, Hottovy and Stechmann [2015] discusses deterministic and stochastic triggers in threshold diffusion models for rainfall and convection. For other applications of the specific threshold diffusion in (1.1) we refer to Lejay and Pigato [2020].

Outline. We state our main results in Section 2 and we discuss them in Section 2.4. We present some numerical studies in Section 3. We collect proofs and technical material in Section 4.

2 Estimators and convergence results

Without loss of generality, we assume the threshold is at $r = 0$. Indeed, given $(X_t)_t$ solution to (1.1) with known level (threshold) $r \in \mathbb{R}$, the process $(Y_t)_t := (X_t - r)_t$ solves (1.1) with threshold at 0 and initial condition $Y_0 = X_0 - r$. As a consequence, all the results in this document can be easily rewritten in the case $r \neq 0$.

So, let $X = (X_t)_{t \in [0, \infty)}$ be the solution to (1.1) with $r = 0$, W Brownian motion, and X_0 deterministic (see Le Gall [1985] or Bass and Chen [2005] for strong existence results). We assume that $b_- > 0$ and $b_+ < 0$, so that the process is mean reverting and ergodic. In this case, also the discrete time process $(X_{kh})_{k \in \mathbb{N}}$ (the so-called h -skeleton sampled chain) is ergodic, for any $h > 0$ (this can be proved as in [Hu and Xi, 2022, Lemma 1]).

Given $T \in (0, \infty)$, we assume to observe X over a time grid with h discretization step such that $T = hN$. The (symmetric) local time of X at 0, denoted by $L_T(X)$, is a continuous process quantifying the amount of time spent by X close to level 0 up to time T . An estimator of the local time $L_T(X)$ is given by

$$L_{h,N} := 2 \sum_{k=0}^{N-1} \mathbf{1}_{\{X_{kh} X_{(k+1)h} < 0\}} |X_{(k+1)h}|. \quad (2.1)$$

We refer to Section 2.3 for details on the local time, convergence results and alternative local time estimators. Writing $\pm \in \{-, +\}$, let us introduce quantities

$$M_T^\pm := \int_0^T \mathbf{1}_{\{\pm X_s \geq 0\}} dX_s \quad \text{and} \quad Q_T^\pm := \int_0^T \mathbf{1}_{\{\pm X_s \geq 0\}} ds \quad (2.2)$$

and their discrete counterparts

$$M_{h,N}^\pm := \sum_{k=0}^{N-1} \mathbf{1}_{\{\pm X_{kh} \geq 0\}} (X_{(k+1)h} - X_{kh}) \quad \text{and} \quad Q_{h,N}^\pm := h \sum_{k=0}^{N-1} \mathbf{1}_{\{\pm X_{kh} \geq 0\}}. \quad (2.3)$$

Throughout the paper we use the notation $\mathcal{N}(\mu, \Sigma)$ for a Gaussian variable with mean vector μ and covariance matrix Σ . We also write $\mathcal{N} = \mathcal{N}(0, \text{Id})$ for a standard Gaussian. We use in what follows the notion of *stable convergence*, that we denote $\xrightarrow{\text{stably}}$, for which we refer to Rényi [1963], Jacod and Shiryaev [2003], Jacod and Protter [2012]. We also write $\xrightarrow{\text{a.s.}}$ for almost sure convergence, $\xrightarrow{\mathbf{P}}$ for convergence in probability, $\xrightarrow{\text{law}}$ for convergence in law.

2.1 Drift parameters inference

We consider estimators for the drift parameters b_+, b_- in (1.1), based on the quantities above (see next Section 2.4 for a derivation). The first one is a GME

$$\left(\bar{b}_{h,N}^+, \bar{b}_{h,N}^-\right) = \left(-\frac{L_{h,N}}{2Q_{h,N}^+}, \frac{L_{h,N}}{2Q_{h,N}^-}\right). \quad (2.4)$$

The second one is a dMLE, that can also be interpreted as a least squares estimator (LSE)

$$\left(\widehat{b}_{h,N}^+, \widehat{b}_{h,N}^-\right) = \left(\frac{M_{h,N}^+}{Q_{h,N}^+}, \frac{M_{h,N}^-}{Q_{h,N}^-}\right). \quad (2.5)$$

In what follows, we provide a complete asymptotic theory for these estimators. We consider different settings, depending on the assumptions on h and T , which can be as follows:

1. h is a fixed constant and $N \rightarrow \infty$ (*long time*, since $T = Nh$, and *low frequency*);
2. $h = h_N = T/N$, with T fixed and $N \rightarrow \infty$ (*high frequency*);
3. $h = h_N = T_N/N$, with $T_N \rightarrow \infty$ and $h_N \rightarrow 0$ as $N \rightarrow \infty$ (*long time and high frequency*).

We show that the estimators in (2.4) and (2.5) are *asymptotically equivalent* as the time horizon goes to infinity (see Lemma 3 in Section 4.2), so that both next Theorems 1 and 2 hold for both estimators, in exactly the same form. For this reason, we state our results in extended form only for \bar{b}^\pm .

We first consider the long time, low frequency framework, meaning that we observe the process over a discrete time grid with fixed time lag, with number of observations going to infinity.

Theorem 1. *Let $h > 0$ fixed. Then*

i) *the estimator is strongly consistent: $\bar{b}_{h,N}^\pm \xrightarrow[N \rightarrow \infty]{\text{a.s.}} b_\pm$,*

ii) *the estimator is asymptotically normal, i.e.*

$$\sqrt{N} \begin{pmatrix} \bar{b}_{h,N}^+ - b_+ \\ \bar{b}_{h,N}^- - b_- \end{pmatrix} \xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, \Gamma),$$

where Γ is given in (4.26) in Lemma 7.

These results hold also substituting estimator \bar{b}^\pm with \widehat{b}^\pm , but in this case the convergence in i) holds in probability (weak consistency).

The result for \bar{b}^\pm is proved in Section 4.3 using ergodic theorems. By Lemma 3 in Section 4.2, we deduce the result for \widehat{b}^\pm from the one for \bar{b}^\pm .

We now assume to observe the process over a discrete time grid with vanishing time lag, with the number of observations going to infinity in a way so that the time horizon also goes to infinity (long time and high frequency framework).

Theorem 2. *Let $(T_N)_{N \in \mathbb{N}}, (h_N)_{N \in \mathbb{N}}$ be positive sequences satisfying $h_N = T_N/N$ and*

$$\lim_{N \rightarrow \infty} T_N = \infty \quad \text{and} \quad \lim_{N \rightarrow \infty} h_N = 0.$$

Then

i) the estimator is consistent: $\bar{b}_{h_N, N}^\pm \xrightarrow[N \rightarrow \infty]{\mathbf{P}} b_\pm$,

ii) if in addition

$$\lim_{N \rightarrow \infty} T_N h_N = 0, \quad (2.6)$$

the estimator is asymptotically normal, i.e.

$$\sqrt{T_N} \begin{pmatrix} \bar{b}_{h_N, N}^+ - b_+ \\ \bar{b}_{h_N, N}^- - b_- \end{pmatrix} \xrightarrow[N \rightarrow \infty]{\text{stably}} \mathcal{N}(0, \Sigma)$$

where

$$\Sigma := \frac{b_- + |b_+|}{b_- |b_+|} \begin{pmatrix} \sigma_+^2 |b_+| & 0 \\ 0 & \sigma_-^2 b_- \end{pmatrix}.$$

Precisely the same results hold substituting estimator \bar{b}^\pm with \hat{b}^\pm .

The result for \hat{b}^\pm is proved in Section 4.4. The result for \bar{b}^\pm follows from Lemma 3 in Section 4.2.

The next result shows that when the time horizon is fixed, both the estimators above, based on discrete observations, converge in high frequency, with speed $N^{1/4}$, to their continuous time analogues.

Theorem 3. Let $T \in (0, \infty)$ be fixed and $h_N = T/N$. Then, both

$$N^{1/4} \left(\hat{b}_{h_N, N}^+ - \frac{M_T^+}{Q_T^+}, \hat{b}_{h_N, N}^- - \frac{M_T^-}{Q_T^-} \right) \quad \text{and} \quad N^{1/4} \left(\bar{b}_{h_N, N}^+ - \frac{-L_T}{2Q_T^+}, \bar{b}_{h_N, N}^- - \frac{L_T}{2Q_T^-} \right), \quad (2.7)$$

when $N \rightarrow \infty$, converge stably to

$$\left(\frac{1}{Q_T^+}, -\frac{1}{Q_T^-} \right) \sqrt{\frac{4\sqrt{T}}{3\sqrt{2\pi}} \frac{\sigma_-^2 + \sigma_+^2}{\sigma_- + \sigma_+}} \sqrt{L_T(X)} \mathcal{N} \quad (2.8)$$

with \mathcal{N} a standard Gaussian random variable independent of X , and $L_T(X)$ local time of X at 0, up to time T .

This theorem is proved in Section 4.2.1.

2.2 Volatility parameters inference

We now introduce an estimator for the volatility parameters σ_\pm based on discrete observations with a fixed h discretization step. Let us denote

$$Q_{h, N}^{\pm, 1} := h \sum_{k=0}^{N-1} X_{kh} \mathbf{1}_{\{\pm X_{kh} \geq 0\}} \quad (2.9)$$

and

$$(\bar{\sigma}_{h, N}^\pm)^2 = \pm \frac{1}{2} \frac{L_{h, N} Q_{h, N}^{\pm, 1}}{(Q_{h, N}^\pm)^2}. \quad (2.10)$$

This is a GME estimator as the one in (2.4). We consider its long time behavior.

Theorem 4. Let $h > 0$ fixed. Then

i) the estimator is consistent: $\bar{\sigma}_{h,N}^{\pm} \xrightarrow[N \rightarrow \infty]{\mathbf{P}} \sigma_{\pm}$,

ii) the estimator is asymptotically normal, i.e.

$$\sqrt{N} \begin{pmatrix} (\bar{\sigma}_{h,N}^+)^2 - \sigma_+^2 \\ (\bar{\sigma}_{h,N}^-)^2 - \sigma_-^2 \end{pmatrix} \xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, \Gamma_{\sigma})$$

where Γ_{σ} is given in (4.28).

This theorem is proved in Section 4.3.

When high frequency observations are available, the quadratic variation estimators considered in Lejay and Pigato [2018] seem more suitable for estimating σ_{\pm} . However, $\bar{\sigma}_{h,N}^{\pm}$ in (2.10) could be the better choice for estimating σ_{\pm} when X is observed at a low frequency, since high frequency observations are not required for convergence. Let us also note that a joint version of Theorems 1 and 4 could be written, providing a joint convergence of $\bar{b}_{h,N}^{\pm}, \bar{\sigma}_{h,N}^{\pm}$, with speed \sqrt{N} and asymptotic covariance again of the form in (4.15), this time a 4 by 4 matrix.

2.3 Local time estimation

Given a semi-martingale $Y : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ and $z \in \mathbb{R}$, $T \in [0, \infty)$, the quantity

$$L_T^z(Y) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^T \mathbf{1}_{\{-\epsilon \leq Y_s - z \leq \epsilon\}} d\langle Y \rangle_s$$

defines the (symmetric) local time of Y at z , which quantifies the amount of time spent by Y close to level z up to time T , properly re-scaled. We focus here on the local time at the discontinuity of X , that is $L_T := L_T^0(X)$. Because of the ergodic property, the rescaled local time converges:

$$\frac{L_T(X)}{T} \xrightarrow[T \rightarrow \infty]{\text{a.s.}} \frac{2|b_+|b_-}{b_- + |b_+|} =: L_{\infty} \quad (2.11)$$

(cf. [Borodin and Salminen, 2015, Chapter II 35(c)]). There are several ways to approximate the local time L_T from discrete observations of X . We have already introduced the estimator $L_{h,N}$ in (2.1). A well known estimator is the one from renormalization of number of crossings, which is

$$\bar{L}_{h,N} = \sqrt{\frac{\pi}{2}} \frac{\sigma_+ + \sigma_-}{2} \sqrt{h} \sum_{k=0}^{N-1} \mathbf{1}_{\{X_{kh} X_{(k+1)h} < 0\}}.$$

We define also the estimator “from sample covariance”

$$\hat{L}_{h,N} = -\frac{3\sqrt{\pi}}{2\sqrt{2}} \frac{\sigma_- + \sigma_+}{\sigma_- \sigma_+} \frac{1}{\sqrt{h}} \sum_{k=0}^{N-1} [X^+, X^-]_N^h$$

where $[Y, Z]_N^h = \sum_{k=0}^{N-1} (Y_{(k+1)h} - Y_{kh})(Z_{(k+1)h} - Z_{kh})$ is the discrete covariation of two one-dimensional processes Y, Z , over N observations with a time lag h . The process $X^+ = X \mathbf{1}_{\{X \geq 0\}}$ denotes the positive part of X and $X^- = -X \mathbf{1}_{\{X < 0\}}$ its negative part (which is positive).

The three statistics above converge to the local time for fixed final horizon $T = h_N N$ when N (the number of observations) goes to infinity, with speed $N^{1/4}$ (see Mazzonetto [2019]). We now study their behavior for low frequency (fixed time lag) observations, in long time.

Let us denote with $O(h)$ any function $f(h)$ s.t. $f(h)/h \rightarrow C \in \mathbb{R}$, as $h \rightarrow 0$.

Lemma 1. Let $L_{h,N}, \bar{L}_{h,N}, \hat{L}_{h,N}$ be the local time estimators defined above and in (2.1). Then,

$$\frac{L_{h,N}}{hN} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} L_\infty \quad (2.12)$$

$$\frac{\bar{L}_{h,N}}{hN} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \bar{L}(h) = L_\infty + l_1\sqrt{h} + O(h) \quad (2.13)$$

$$\frac{\hat{L}_{h,N}}{hN} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \hat{L}(h) = L_\infty + l_2\sqrt{h} + O(h) \quad (2.14)$$

where L_∞ is in (2.11), $\bar{L}(h), \hat{L}(h)$ are functions of the parameters and of h ,

$$l_1 = \sqrt{\frac{\pi}{2}} \frac{|b_+|b_-}{\sigma_+ + \sigma_-} > 0 \text{ and } l_2 = -\frac{3\sqrt{\pi}}{4\sqrt{2}} \frac{|b_+|b_-}{|b_+| + b_-} \left(\frac{|b_+|}{\sigma_+} + \frac{b_-}{\sigma_-} + \frac{|b_+| + b_-}{\sigma_- + \sigma_+} \right) < 0.$$

Moreover,

$$\begin{aligned} \sqrt{N} \left(\frac{L_{h,N}}{hN} - L_\infty \right) &\xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, \Gamma_l), \\ \sqrt{N} \left(\frac{\bar{L}_{h,N}}{hN} - \bar{L}(h) \right) &\xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, \Gamma_{\bar{l}}), \\ \sqrt{N} \left(\frac{\hat{L}_{h,N}}{hN} - \hat{L}(h) \right) &\xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, \Gamma_{\hat{l}}), \end{aligned}$$

where $\Gamma_l, \Gamma_{\bar{l}}, \Gamma_{\hat{l}}$ can be computed from (4.15).

This lemma is proved in Section 4.3. The behavior of the three estimators is consistent with (2.11) if we only look at the first term in the expansion of the limit. However, the \sqrt{h} term is non zero for the limit of $\bar{L}_{h,N}$ and $\hat{L}_{h,N}$. So the only estimator having the same limit behavior as the local time, when observations are in low frequency and the time horizon goes to infinity, is $L_{h,N}$. Furthermore, note that $L_{h,N}$ is the only one that does not require previous knowledge of the volatility parameters.

2.4 Derivation of the estimators and comments on the results

Derivation of the estimators. The dMLE in (2.5) has been proposed in [Lejay and Pigato, 2020] and is the maximum of a discretized classical likelihood for the solution of a SDE, based on the Girsanov transform. It can also be derived as a LSE with contrast function

$$\sum_{k=0}^{N-1} \left| X_{(k+1)h} - X_{kh} - hb^+ \mathbf{1}_{\{X_{kh} \geq 0\}} - hb^- \mathbf{1}_{\{X_{kh} < 0\}} \right|^2.$$

The GME (2.4) is related to the one proposed for a different threshold SDE in [Hu and Xi, 2022]. If the process is ergodic, the GME estimator is derived from the inversion of quantities such as $\mathbf{E}[(X_\infty)^p \mathbf{1}_{\{\pm X_\infty > 0\}}]$, where $p \in \mathbb{N}$ and X_∞ is a r.v. whose distribution is the invariant one (it has density (4.2), see Section 4.1). In particular, one can compute a notion of asymptotic occupations time as

$$Q_\infty^\pm := \mathbf{E}[\mathbf{1}_{\{\pm X_\infty \geq 0\}}] = \frac{|b_\mp|}{b_- + |b_+|}, \quad Q_\infty^{\pm,1} := \mathbf{E}[X_\infty \mathbf{1}_{\{\pm X_\infty \geq 0\}}] = \frac{\pm |b_\mp| \sigma_\pm^2}{2|b_\pm|(b_- + |b_+|)}. \quad (2.15)$$

These quantities can also be approximated from discrete observations using the ergodic theorem (cf. next equations (4.16) and (4.19)). From the approximations $L_{h,N}$ to L_∞ (cf. next

equations (2.11)-(2.12)) and $Q_{h,N}^\pm$ to Q_∞^\pm one can derive the estimators $\bar{b}_{h,N}^\pm$ in (2.4). Note that similar estimators could be defined using $\hat{L}_{h,N}$ or $\bar{L}_{h,N}$, but these would not have the correct limit (they would not be consistent) as $N \rightarrow \infty$, for fixed $h > 0$, because of Lemma 1. However, for small h they could give good approximations.

Similarly, the definition of the volatility estimator in (2.10) follows from the same approximations and the asymptotic value $Q_\infty^{\pm,1}$ of $Q_{h,N}^{\pm,1}$, corresponding to the first conditional moment used in Hu and Xi [2022]. Note the dependence on σ_\pm in $Q_\infty^{\pm,1}$, not present in Q_∞^\pm and L_∞ .

The local time estimators are special cases of the class of statistics considered for instance in Jacod [1998] for Brownian diffusions for fixed time horizon. In the context of threshold diffusions for fixed time horizon, $L_{h,N}$ was considered in Lejay and Pigato [2020], the estimator related to the number of crossings, $\bar{L}_{h,N}$, was considered in Portenko [1994], Lejay et al. [2019], and $\hat{L}_{h,N}$ was considered in Lejay and Pigato [2018].

On the equivalence of drift GME and dMLE, beyond the ergodic regime. In Lemma 3 we see that, in the ergodic case, GME and dMLE are equivalent in long time. This is because the dMLE is essentially given by the sum of two parts, one corresponding to the GME, the other given by the final and initial value of the process at T , normalized with the occupation time:

$$\widehat{b}_{h,N}^\pm = \bar{b}_{h,N}^\pm \pm \frac{\{X_{hN}\}^\pm - \{X_0\}^\pm}{Q_{h,N}^\pm} \quad \mathbf{P}\text{-a.s.} \quad (2.16)$$

This equation is proved in the proof of Lemma 3 in Section 4.2. In the ergodic case, the fraction in (2.16) vanishes as $T \rightarrow \infty$ (see Lemma 2 in Section 4.2).

A complete statistical analysis of the estimators from continuous time observations is proposed in [Lejay and Pigato, 2020, Propositions 3-7, 10-12], more precisely the estimators behavior is studied even if the process is not ergodic. The results of the present document imply the same results for the dMLE in high frequency and infinite horizon. In this paper we deal only with the ergodic case in Theorem 2). The proof, for non ergodic cases, is the same as the one provided in Section 4.4 since Lemmas 9-10 also hold for non-ergodic cases. In the null recurrent case with non-vanishing drift, one needs to impose that $h_N T_N^{3/2} \rightarrow 0$.

When the process is not ergodic, GME is not necessarily equivalent to the dMLE. For instance, in long time, in the transient case, the fraction in (2.16) does not vanish on the side where the process stays indefinitely, and it actually is the term providing the convergence of the estimators to the drift parameter. This is reminiscent of the drift MLE in a simple drifted Brownian motion (in finance, the Black-Scholes model), which does not depend on the intermediate observations, but only on the process value at T and therefore, does not depend on the frequency of the observations. Note that if the drift is not 0, the Black-Scholes model is transient, while the model in (1.1) can be transient or recurrent/ergodic depending on the drift parameters. The normalized final value of the process at time T is also the leading term for the model (1.1), when at least one of the drift parameters is such that the process is not mean reverting, which includes some null-recurrent cases.

Drift estimation in high frequency and infinite horizon. Condition (2.6) is in line with what has been obtained in the literature so far for non-threshold diffusions, and it is the best condition obtained so far for threshold ones. Up to the authors knowledge, Theorem 2 is the first result for threshold diffusions allowing for deterministic initial conditions (instead of starting at the stationary distributions) when looking at high-frequency observations over an infinite time horizon.

Theorem 2 holds also for non-equally spaced high frequency observations. More precisely, the process is observed at times $0 = t_0 < t_1 < \dots < t_N = T_N$ and h_N represents the maximal

observation lag: $h_N = \sup_{k=0, \dots, N-1} (t_{k+1} - t_k)$.

Multi-threshold. The method we are presenting here should also work in a multi-threshold setting, at least in the high frequency and long time setting, where it works analogously to [Mazonetto and Pigato, 2024, Supplementary material Section S.3]. In the GME case, with more than one threshold, the definition of the estimators would involve more complicated, implicit expressions, that one would need to solve using numerical methods, while, if there is only one threshold, explicit expressions are available. Therefore, for the sake of simplicity of the exposition and ease of implementation, we stick here to the case with only one threshold.

3 Numerical studies

3.1 Comparison of drift estimators

We compare here estimators $\bar{b}_{h,N}^\pm$ with $\hat{b}_{h,N}^\pm$, with $h > 0$ fixed, as the number of observations goes to infinity, plotting the mean squared error of the estimators (MSE), for a process with parameters as in Table 1, simulated via the Euler-Maruyama scheme. Clearly, because of The-

b_-	b_+	σ_-	σ_+
0.02	-0.01	0.07	0.10

Table 1: Simulations parameters.

orems 1 and 2, the asymptotic behavior of the estimators is the same. However, our numerical experiments suggest that the GME slightly outperforms the dMLE, not by a significant margin, but consistently over different sets of parameters and time horizons (see Figure 1). This can be explained by the discussion in Section 2.4, since we are in the ergodic setting.

3.2 The asymptotic normality constant in the drift estimator

We look now at the CLT in Theorem 1 - ii). The limit covariance matrix is given in (4.26), for which we do not have an analytical expression. However, it can be evaluated numerically, on synthetic data, simulating several realizations of the process and then computing the empirical covariance. We do so, and plot in Figure 2 the empirical distribution of the error, rescaled with \sqrt{N} , for various lengths of the observed time series, generated with parameters as in Table 1 via the Euler-Maruyama scheme. A similar numerical representation for Theorem 2 can be found in Lejay and Pigato [2020], where the continuous time equivalent of Theorem 2 is stated and checked numerically.

In view of the independence of the limit Gaussians in Theorem 2, an interesting question is whether the limit cross covariance vanishes or not also in Theorem 1. In our simulations, we can never reject the null hypothesis of 0 limit cross covariance. However, this numerical evidence is not conclusive since the limit cross covariance could be small but non-zero, and still be consistent with our simulations.

These simulations provide a numerical validation of Theorem 1. However, this approach generally cannot be used to estimate the error when dealing with parameters inference from empirical data, since in this case estimation can only be based on one observed time series and several realizations are not available. Estimating the limit covariance can be necessary, for example in order to perform a test on the estimated parameters. In this case, one way to estimate the limit covariance is to split the time series in several batches and use them as separate realizations, as explained in [Brooks et al., 2011, Section 1.10.1]. Alternative methods

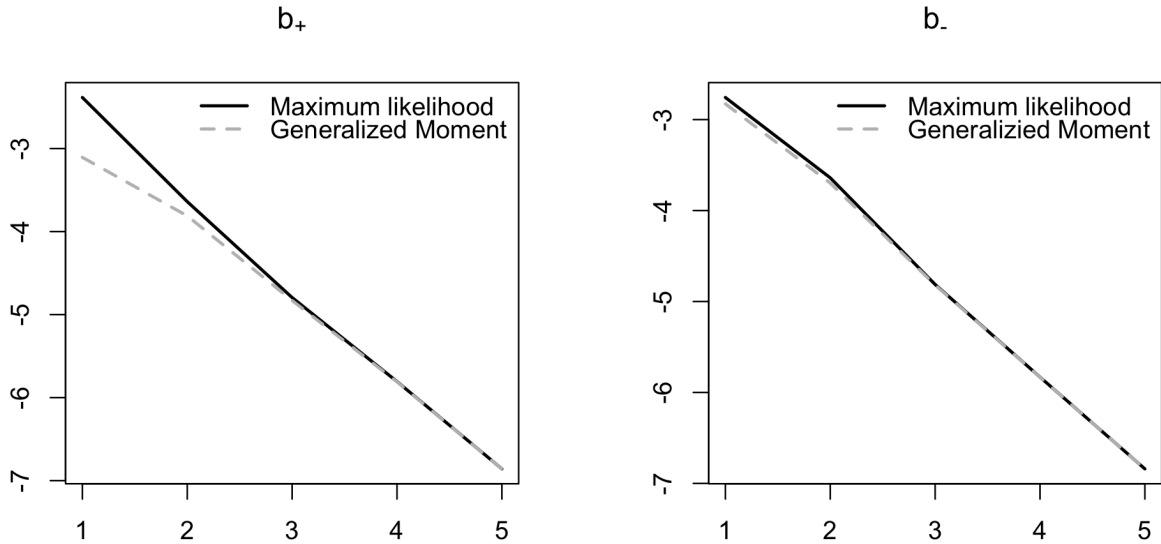


Figure 1: We compare the GME $\bar{b}_{h,N}^\pm$ and the dMLE $\hat{b}_{h,N}^\pm$, with $h = 1$ fixed, varying the time horizon (length of the time series) N . We plot in this figure the mean squared error of each estimator, on 10^3 simulated time series, as a function of the length of the time series, in a log-log plot in base 10.

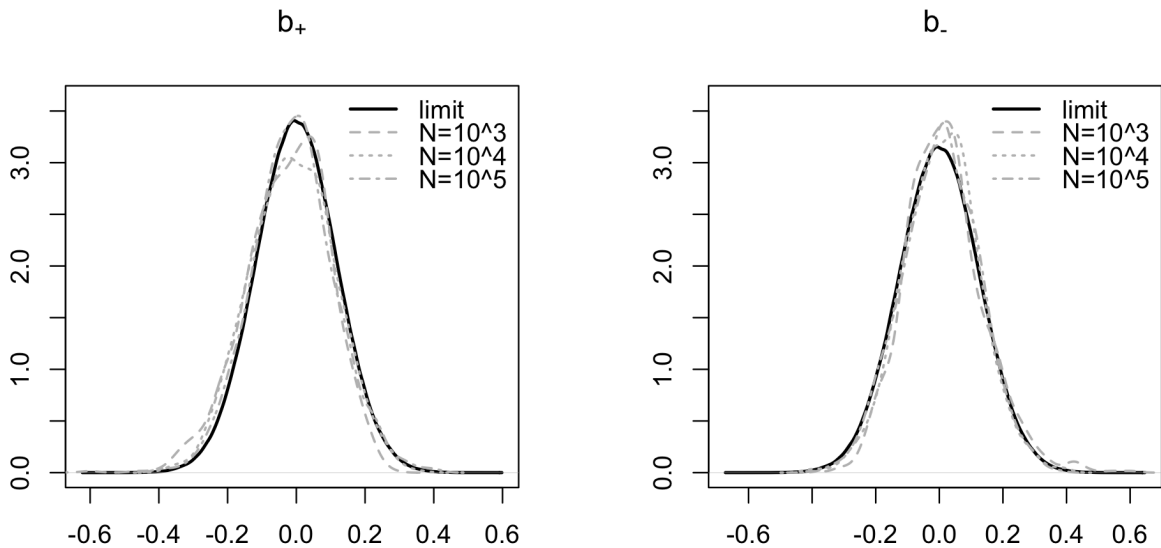


Figure 2: CLT in Theorem 1-ii), with parameters as in Table 1. We plot the density of the theoretical distribution of the estimation error, with variance estimated on 10^3 simulated paths, and compare it with the distribution of the rescaled error on $n = 10^3$ paths, and $N = 10^3, 10^4, 10^5$ observations on each path.

are presented e.g. in [Brooks et al., 2011, Section 1.10.2 and 1.10.3], see also [Brooks et al., 2011, Section 1.12.9] for subsampling techniques.

4 Technical material and proofs

We collect in this section all the mathematical proofs. In Section 4.1 we give some preliminary technical results related to the law of the process via its infinitesimal generator. In Section 4.2 we prove the asymptotic equivalence of the GME and dMLE and Theorem 3. In Section 4.3, we provide the proofs related to the GME, with fixed time lag and long time horizon. Finally, in Section 4.4 the proofs related to the dMLE, in high frequency and with long time horizon.

4.1 Technical tools: scale function, speed measure, fundamental system and resolvent kernel

The infinitesimal generator \mathcal{A} of the process X solution to (1.1) can be written as

$$\mathcal{A}f = \frac{1}{2}\sigma^2(x)e^{-h(x)}\frac{d}{dx}\left(e^{h(x)}\frac{df(x)}{dx}\right) \quad \text{with } h(x) = \int_0^x \frac{2b(y)}{\sigma^2(y)} dy,$$

cf. e.g. [Itô and McKean, 1996] for a general discussion and [Lejay and Pigato, 2020] for the diffusion in (1.1). The process X is a diffusion (roughly-speaking, a strong Markov process with continuous sample paths) which is fully characterized by its *speed measure* M with density m and *scale function* S given by

$$m(x) := \frac{2}{\sigma(x)^2} \exp(h(x)) \quad \text{and} \quad S(x) := \int_0^x \exp(-h(y)) dy. \quad (4.1)$$

When the process is positive or null recurrent, the speed measure is a stationary measure (invariant measure for the transition semigroup). When $b_+ < 0$, $b_- > 0$ we say that the process is *ergodic* and its *invariant distribution* is the re-normalized speed measure, denoted by μ , with density given by

$$\frac{m(x)}{M(\mathbb{R})} = \begin{cases} \frac{2}{\sigma_+^2} \times \frac{|b_+|b_-}{b_- + |b_+|} e^{\frac{-2|b_+|x}{\sigma_+^2}} & \text{if } x \geq 0, \\ \frac{2}{\sigma_-^2} \times \frac{b_-|b_+|}{b_- + |b_+|} e^{\frac{2b_-x}{\sigma_-^2}} & \text{if } x < 0. \end{cases} \quad (4.2)$$

Here $M(\mathbb{R}) = \int_{\mathbb{R}} m(x)dx = (|b_+| + b_-)/(|b_+|b_-)$. (Note that there is a typo in [Lejay and Pigato, 2020].) Throughout the paper, we write X_∞ for a r.v. independent of X , following the stationary distribution, i.e., with density (4.2). Let us consider equation

$$(\mathcal{A} - \lambda)u = 0. \quad (4.3)$$

For fixed $\lambda > 0$, the set of solutions to (4.3) is a two-dimensional vector space. There exist two continuous, positive functions $\psi_\lambda, \varphi_\lambda$ solution to (4.3), with ψ_λ increasing from 0 to ∞ and φ_λ decreasing from ∞ to 0, such that $\psi_\lambda(0) = \varphi_\lambda(0) = 1$, called the minimal functions. Such functions can be taken as a basis for the space of solutions to (4.3). With coefficients as in (1), the minimal functions are explicit and given as

$$\psi_\lambda(x) = \begin{cases} \exp\left(x \frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2\lambda}}{\sigma_-^2}\right) & \text{if } x < 0 \\ \kappa_+ \exp\left(x \frac{-b_+ + \sqrt{b_+^2 + 2\sigma_+^2\lambda}}{\sigma_+^2}\right) + \delta_+ \exp\left(x \frac{-b_+ - \sqrt{b_+^2 + 2\sigma_+^2\lambda}}{\sigma_+^2}\right) & \text{if } x \geq 0, \end{cases} \quad (4.4)$$

$$\varphi_\lambda(x) = \begin{cases} \kappa_- \exp\left(x \frac{-b_- - \sqrt{b_-^2 + 2\sigma_-^2\lambda}}{\sigma_-^2}\right) + \delta_- \exp\left(x \frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2\lambda}}{\sigma_-^2}\right) & \text{if } x < 0, \\ \exp\left(x \frac{-b_+ - \sqrt{b_+^2 + 2\sigma_+^2\lambda}}{\sigma_+^2}\right) & \text{if } x \geq 0 \end{cases} \quad (4.5)$$

with

$$\begin{aligned}\kappa_+ &:= \frac{-b_- \sigma_+^2 + b_+ \sigma_-^2 + \sigma_-^2 \sqrt{b_+^2 + 2\lambda \sigma_+^2} + \sigma_+^2 \sqrt{b_-^2 + 2\lambda \sigma_-^2}}{2\sigma_-^2 \sqrt{b_+^2 + 2\lambda \sigma_+^2}}, \\ \delta_+ &:= \frac{b_- \sigma_+^2 - b_+ \sigma_-^2 + \sigma_-^2 \sqrt{b_+^2 + 2\lambda \sigma_+^2} - \sigma_+^2 \sqrt{b_-^2 + 2\lambda \sigma_-^2}}{2\sigma_-^2 \sqrt{b_+^2 + 2\lambda \sigma_+^2}}, \\ \kappa_- &:= \frac{-b_- \sigma_+^2 + b_+ \sigma_-^2 - \sigma_-^2 \sqrt{b_+^2 + 2\lambda \sigma_+^2} + \sigma_+^2 \sqrt{b_-^2 + 2\lambda \sigma_-^2}}{2\sigma_+^2 \sqrt{b_-^2 + 2\lambda \sigma_-^2}}, \\ \delta_- &:= \frac{b_- \sigma_+^2 - b_+ \sigma_-^2 + \sigma_-^2 \sqrt{b_+^2 + 2\lambda \sigma_+^2} + \sigma_+^2 \sqrt{b_-^2 + 2\lambda \sigma_-^2}}{2\sigma_+^2 \sqrt{b_-^2 + 2\lambda \sigma_-^2}}.\end{aligned}$$

Using the minimal functions, one can define the Wronskian of the diffusion as

$$\begin{aligned}W_\lambda &= \varphi_\lambda(x) \frac{\partial_x \psi_\lambda(x)}{\partial_x S(x)} - \psi_\lambda(x) \frac{\partial_x \varphi_\lambda(x)}{\partial_x S(x)} \\ &= \frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2} - \frac{-b_+ - \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2} \\ &= \frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2} + \frac{b_+ + \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2}\end{aligned}$$

where S is the scale function in (4.1). Note that

$$W_\lambda = \lambda \frac{b_- + |b_+|}{b_- |b_+|} + O(\lambda^2) \quad (4.6)$$

Let us define the *resolvent kernel* of (1.1) as the Laplace transform of the transition density

$$r(\lambda, x, y) = \mathcal{L}_\lambda(p(\cdot, x, y)) = \int_0^\infty p(t, x, y) \exp(-\lambda t) dt. \quad (4.7)$$

It can be computed using the speed measure (4.1), the minimal functions (4.4), (4.5) and the Wronskian, as

$$r(\lambda, x, y) = \frac{m(y)}{W_\lambda} \begin{cases} \psi_\lambda(x) \varphi_\lambda(y) & \text{if } x < y, \\ \varphi_\lambda(x) \psi_\lambda(y) & \text{if } x > y. \end{cases} \quad (4.8)$$

We can use it as follows to compute Laplace transforms of expectations of functions $f(\cdot)$ of X_t starting from a point z , if we can exchange the order of time integral and expectation, since

$$R_\lambda f(z) := \mathcal{L}_\lambda \mathbf{E}_z[f(z, X_\cdot)] = \int_0^\infty \mathbf{E}_z[e^{-\lambda t} f(z, X_t)] dt = \int_{\mathbb{R}} r(\lambda, z, y) f(z, y) dy. \quad (4.9)$$

4.2 The relation between different estimators

Lemma 2. *Let X be solution to (1.1). Then*

$$t^{-p} X_t \xrightarrow[t \rightarrow \infty]{\mathbf{P}} 0, \text{ for any } p \in (0, +\infty). \quad (4.10)$$

Proof. Step 1: We first prove that the statement holds with initial condition $X_0 = 0$. We show that

$$\mathbf{E}[|X_t|] \rightarrow c \in \mathbb{R}, \text{ for } t \rightarrow \infty \quad (4.11)$$

which implies the statement. We use that $\mathbf{E}[|X_t|] = \mathbf{E}[X_t^+] + \mathbf{E}[X_t^-]$ and write

$$\mathcal{L}_\lambda \mathbf{E}[X_t^+] = \int_0^\infty r(\lambda, 0, y) y dy$$

with r in (4.7). Using (4.8), we compute

$$\mathcal{L}_\lambda \mathbf{E}[X_t^+] = \frac{\psi_\lambda(0)}{W_\lambda} \int_0^\infty m(y) \varphi_\lambda(y) y dy = \frac{1}{W_\lambda} \int_0^\infty m(y) \varphi_\lambda(y) y dy.$$

We have

$$\begin{aligned} \int_0^\infty m(y) \varphi_\lambda(y) y dy &= \frac{2}{\sigma_+^2} \int_0^\infty \exp\left(y \frac{b_+ - \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2}\right) y dy \\ &= \frac{2\sigma_+^2}{(b_+ - \sqrt{b_+^2 + 2\sigma_+^2 \lambda})^2} = \frac{(\sqrt{b_+^2 + 2\sigma_+^2 \lambda} + b_+)^2}{2\sigma_+^2 \lambda^2} \end{aligned}$$

and, using also (4.6), as $\lambda \rightarrow 0$,

$$\mathcal{L}_\lambda \mathbf{E}[X_t^+] = \frac{\sigma_+^2 b_-}{2|b_+|(b_- + |b_+|)} \frac{1}{\lambda} + O(1).$$

So, using the final value theorem (Tauberian theorem), we see that as $t \rightarrow \infty$

$$\mathbf{E}[X_t^+] \rightarrow \frac{\sigma_+^2 b_-}{2|b_+|(b_- + |b_+|)}$$

which is $\mathbf{E}[X_\infty^+]$ from (4.2). With a similar computation we prove an analogous result for $\mathbf{E}[X_t^-]$, and therefore as $t \rightarrow \infty$ we obtain $\mathbf{E}[X_t] \rightarrow \mathbf{E}[X_\infty]$ and $\mathbf{E}[|X_t|] \rightarrow \mathbf{E}[|X_\infty|]$, where the limits can be directly computed and are finite. Markov's inequality completes the proof of this first step.

Step 2: The statement for $X_0 = 0$ implies the statement for any $X_0 \in \mathbb{R}$, say $X_0 = x$. Let us denote the process with such initial condition by X . Let τ be the stopping time identifying the first time for which $X_t = 0$. Note that τ is a.s. finite since X is recurrent. The process $Y_t := X_{t+\tau}$ is a weak solution to (1.1) with $Y_0 = 0$. Hence, by the strong Markov property, we deduce that $\mathbf{E}[|Y_t|]$ converges to a constant as t goes to infinity. To prove the statement it suffices to show that $\sup_{t>0} \mathbf{E}[|Y_t - X_t|]$ is bounded by a constant. From (1.1), follows that $X_{t+\tau} - X_t = \int_t^{t+\tau} b(X_s) ds + \int_t^{t+\tau} \sigma(X_s) dW_s$. Then, Hölder's inequality and Itô isometry yield for all $t > 0$

$$\mathbf{E}[|Y_t - X_t|] = \mathbf{E}[|X_{t+\tau} - X_t|] \leq 2 \max\{|b_+|, b_-, \sigma_-, \sigma_+\} (1 + \mathbf{E}[\tau]),$$

and $\mathbf{E}[\tau] < \infty$ because the process is positive recurrent. The proof is thus completed. \square

Lemma 3. For $h > 0$ fixed

$$\sqrt{N} \left(\bar{b}_{h,N}^\pm - \hat{b}_{h,N}^\pm \right) \xrightarrow[N \rightarrow \infty]{\mathbf{P}} 0.$$

For $h_N = T_N/N$, with T_N and N satisfying condition (2.6) and such that $\lim_{N \rightarrow \infty} T_N = \infty$,

$$\sqrt{h_N N} \left(\bar{b}_{h_N, N}^\pm - \hat{b}_{h_N, N}^\pm \right) \xrightarrow[N \rightarrow \infty]{\mathbf{P}} 0.$$

For $T > 0$ fixed, $h_N = T/N$ as $N \rightarrow \infty$,

$$\bar{b}_{h_N, N}^\pm - \hat{b}_{h_N, N}^\pm \xrightarrow[N \rightarrow \infty]{\mathbf{P}} \mp \frac{\{X_T\}^\pm - \{X_0\}^\pm}{Q_T^\pm}.$$

Proof. Itô-Tanaka formula establishes that the following equalities hold \mathbf{P} -a.s.:

$$\{X_T\}^\pm - \{X_0\}^\pm = \pm M_T^\pm + \frac{1}{2}L_T(X), \quad (4.12)$$

Moreover, it is not difficult to see, similarly to [Mazzonetto and Pigato, 2024, (S2.6) in supplementary material], that for any h, N , we have \mathbf{P} -a.s.

$$\{X_{hN}\}^\pm - \{X_0\}^\pm = \pm M_{hN}^\pm + \frac{1}{2}L_{hN}. \quad (4.13)$$

Hence, \mathbf{P} -a.s.

$$\sqrt{N} \left(\bar{b}_{h,N}^\pm - \widehat{b}_{h,N}^\pm \right) = \mp \frac{1}{\sqrt{h}} \frac{hN}{\sqrt{hN}} \left(\frac{\{X_{hN}\}^\pm - \{X_0\}^\pm}{Q_{h,N}^\pm} \right).$$

For h fixed, when $N \rightarrow \infty$, we have $Q_{h,N}^\pm/(hN) \xrightarrow{\text{a.s.}} Q_\infty^\pm$, because of (4.16). For $h_N = T_N/N$ and $T_N = h_N N$ such that $\lim_{N \rightarrow \infty} T_N = \infty$ and (2.6), we have $Q_{h_N,N}^\pm/T_N \xrightarrow{\text{a.s.}} Q_\infty^\pm$ when $N \rightarrow \infty$, because of the ergodic theorem and Lemma 9. Lemma 2 implies that $X_t/\sqrt{t} \xrightarrow{\mathbf{P}} 0$ when $t \rightarrow \infty$. The first two statements follow.

Now, when $T > 0$ fixed, $h_N = T/N$, we have $Q_{h_N,N}^\pm \xrightarrow{\text{a.s.}} Q_T^\pm$ when $N \rightarrow \infty$ and the last statement follows. \square

4.2.1 Proof of Theorem 3

The result states that, for fixed time horizon, in high frequency, the drift dMLE from discrete observations converges in probability towards the drift MLE from continuous observations (4.29) and provides the convergence rates. It also states that the generalized moment estimator (2.4), converges for fixed time and in high frequency to its continuous-time analog

$$\bar{\ell}_T := \left(-L_T/(2Q_T^+), \quad L_T/(2Q_T^-) \right).$$

In Lemma 3 it is shown that, for fixed time and in high frequency,

$$\widehat{b}_{h_N,N}^\pm - \bar{b}_{h_N,N}^\pm \xrightarrow[N \rightarrow \infty]{\mathbf{P}} \pm \frac{\{X_T\}^\pm - \{X_0\}^\pm}{Q_T^\pm}$$

and, from (4.12)-(4.13) in the proof of Lemma 3, it holds \mathbf{P} -a.s. that

$$\widehat{b}_{h_N,N}^\pm - \beta_T^\pm + \bar{\ell}_T^\pm - \bar{b}_{h_N,N}^\pm = \pm (\{X_T\}^\pm - \{X_0\}^\pm) \frac{Q_T^\pm - Q_{h_N,N}^\pm}{Q_T^\pm Q_{h_N,N}^\pm}.$$

The speed of convergence of the occupation time is at least \sqrt{N} (see Lejay and Pigato [2018]) which is larger than $N^{1/4}$. Hence, \mathbf{P} -a.s.

$$\bar{b}_{h_N,N}^\pm - \bar{\ell}_T^\pm = \widehat{b}_{h_N,N}^\pm - \beta_T^\pm + O(1/\sqrt{N}).$$

We are thus reduced to prove the result for $\widehat{b}_{h_N,N}^\pm$. Then, (4.12) and (4.13) ensure that \mathbf{P} -a.s.:

$$\widehat{b}_{h_N,N}^\pm - \beta_T^\pm = \mp \frac{L_{h_N,N} - L_T}{2Q_T^\pm} + M_T^\pm \frac{Q_T^\pm - Q_{h_N,T}^\pm}{Q_T^\pm Q_{h_N,T}^\pm}$$

Recall $h_N = T/N$. The consistency, i.e. $(\widehat{b}_{h_N,N}^+ - \beta_T^+, \widehat{b}_{h_N,N}^- - \beta_T^-) \xrightarrow{\mathbf{P}} 0$ when $N \rightarrow \infty$, was already provided in [Lejay and Pigato, 2020, Lemma 1], and the convergence rate is based on the one of the local time approximation [Mazzonetto, 2019, Proposition 2]. The proof is thus completed.

4.3 Proofs of asymptotic behaviors in low frequency and long time: GME in Theorem 1 and Theorem 4 and local time in Lemma 1

Let us define the 2-dimensional discrete time process $(\mathbb{X}_k^h)_{k \in \mathbb{N}} = (X_{hk}, X_{h(k+1)})_{k \in \mathbb{N}}$. In the ergodic case, we set $\mathbb{X}_\infty^h = (X_\infty, X_\infty^h)$, a r.v. that follows its stationary distribution, where X_∞ is a r.v. following the stationary distribution μ of (4.2) and X_∞^h is its evolution over a time h following equation (1.1).

We also define $\tilde{X} = (\tilde{X}_t)_t$ a solution to (1.1) with $\tilde{X}_0 \stackrel{\text{law}}{=} X_\infty$ and $(\tilde{\mathbb{X}}_k^h)_{k \in \mathbb{N}} = (\tilde{X}_{hk}, \tilde{X}_{h(k+1)})_{k \in \mathbb{N}}$.

Lemma 4. *Let $h > 0$ fixed, $\pm b_\pm < 0$.*

I The discrete time process $(X_{hk})_{k \in \mathbb{N}}$ is ergodic, with invariant measure in (4.2). For any measurable function $f : \mathbb{R} \rightarrow \mathbb{R}^n$ s.t. $\mathbf{E}[|f(X_\infty)|] < \infty$,

$$\frac{1}{N} \sum_{k=1}^N f(X_{hk}) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \mathbf{E}[f(X_\infty)].$$

II Moreover, if $|f(x)| \leq C|x|^p$ for some $C_0, C_1, p > 0$ we have that

$$\sqrt{N} \left(\frac{1}{N} \sum_{k=1}^N f(X_{hk}) - \mathbf{E}[f(X_\infty)] \right) \xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, \gamma)$$

where setting $\bar{f}(\cdot) = f(\cdot) - \mathbf{E}[f(X_\infty)]$ we have

$$\gamma = \mathbf{E}[\bar{f}(\tilde{X}_0)\bar{f}^T(\tilde{X}_0)] + 2 \sum_{k=1}^{\infty} \mathbf{E}[\bar{f}(\tilde{X}_0)\bar{f}^T(\tilde{X}_k)]. \quad (4.14)$$

III The process $(\mathbb{X}_k^h)_{k \in \mathbb{N}}$ is ergodic and, for any measurable function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^n$ s.t. $\mathbf{E}[|f(\mathbb{X}_\infty^h)|] < \infty$,

$$\frac{1}{N} \sum_{k=1}^N f(\mathbb{X}_k^h) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \mathbf{E}[f(\mathbb{X}_\infty^h)].$$

IV Moreover, if $|f(x)| \leq C_0 + C_1|x|^p$ for some $C_0, C_1, p > 0$, we have that

$$\sqrt{N} \left(\frac{1}{N} \sum_{k=1}^N f(\mathbb{X}_k^h) - \mathbf{E}[f(\mathbb{X}_\infty^h)] \right) \xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, \Gamma)$$

where setting $\bar{f}(\cdot) = f(\cdot) - \mathbf{E}[f(\mathbb{X}_\infty^h)]$ we have

$$\Gamma = \mathbf{E}[\bar{f}(\tilde{\mathbb{X}}_0^h)\bar{f}^T(\tilde{\mathbb{X}}_0^h)] + \sum_{k=1}^{\infty} \mathbf{E}[\bar{f}(\tilde{\mathbb{X}}_0^h)\bar{f}^T(\tilde{\mathbb{X}}_k^h)] + \mathbf{E}[\bar{f}(\tilde{\mathbb{X}}_k^h)\bar{f}^T(\tilde{\mathbb{X}}_0^h)]. \quad (4.15)$$

Remark 1. The quantities defined in (4.14) and (4.15) are always non-negative, so (4.14) and (4.15) are good definitions of variances in $[0, +\infty)$. Moreover, these are real CLTs with convergence rate \sqrt{N} if γ, Γ are postive definite.

Proof. The proof of I and III is completely analogous to [Hu and Xi, 2022, Theorem 2], the main tool being an application of the ergodic theorem in [Meyn and Tweedie, 2009, Theorem 17.0.1]. Let us now discuss IV. Again we start from [Meyn and Tweedie, 2009, Theorem 17.0.1], which only applies to uni-dimensional f . The fact that this holds for a multidimensional f follows

using the Cramer-Wald theorem as described in [Brooks et al., 2011, Section 1.8]. Clearly it is enough to prove it for $n = 2$. Let (U_1, U_2) be a Gaussian r.v. in \mathbb{R}^2 with covariance

$$\mathbf{E} \begin{pmatrix} \bar{f}_1^2(\tilde{X}_0^h) + 2 \sum_{k=1}^{\infty} [\bar{f}_1(\tilde{X}_0^h) \bar{f}_1(\tilde{X}_k^h)] & \bar{f}_1 \bar{f}_2(\tilde{X}_0^h) + \sum_{k=1}^{\infty} \bar{f}_1(\tilde{X}_0^h) \bar{f}_2(\tilde{X}_k^h) + \bar{f}_2(\tilde{X}_0^h) \bar{f}_1(\tilde{X}_k^h) \\ \bar{f}_1 \bar{f}_2(\tilde{X}_0^h) + \sum_{k=1}^{\infty} \bar{f}_1(\tilde{X}_0^h) \bar{f}_2(\tilde{X}_k^h) + \bar{f}_2(\tilde{X}_0^h) \bar{f}_1(\tilde{X}_k^h) & \bar{f}_2^2(\tilde{X}_0^h) + 2 \sum_{k=1}^{\infty} \bar{f}_2(\tilde{X}_0^h) \bar{f}_2(\tilde{X}_k^h) \end{pmatrix}.$$

For $\lambda_1, \lambda_2 \in \mathbb{R}$ we have, from the univariate version as in [Meyn and Tweedie, 2009, Theorem 17.0.1], after some computations, that in law

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \sum_{k=1}^N (\lambda_1 \bar{f}_1(\mathbb{X}_k^h) + \lambda_2 \bar{f}_2(\mathbb{X}_k^h)) = \lambda_1 U_1 + \lambda_2 U_2,$$

because $\lambda_1 U_1 + \lambda_2 U_2$ is a centered Gaussian with variance given by

$$\begin{aligned} & \lambda_1 \left(\mathbf{E}[\bar{f}_1^2(\tilde{X}_0^h)] + 2 \sum_{k=1}^{\infty} \mathbf{E}[\bar{f}_1(\tilde{X}_0^h) \bar{f}_1(\tilde{X}_k^h)] \right) + \lambda_2 \left(\mathbf{E}[\bar{f}_2^2(\tilde{X}_0^h)] + 2 \sum_{k=1}^{\infty} \mathbf{E}[\bar{f}_2(\tilde{X}_0^h) \bar{f}_2(\tilde{X}_k^h)] \right) \\ & + 2\lambda_1 \lambda_2 \left(\mathbf{E}[\bar{f}_1 \bar{f}_2(\tilde{X}_0^h)] + \sum_{k=1}^{\infty} \left(\mathbf{E}[\bar{f}_1(\tilde{X}_0^h) \bar{f}_2(\tilde{X}_k^h)] + \mathbf{E}[\bar{f}_2(\tilde{X}_0^h) \bar{f}_1(\tilde{X}_k^h)] \right) \right). \end{aligned}$$

Therefore Cramer Wald theorem implies (4.15) in the multivariate sense. The proof of II is analogous, with the addition of the fact that

$$\mathbf{E}[\bar{f}_1(\tilde{X}_0^h) \bar{f}_2(\tilde{X}_k^h)] = \mathbf{E}[\bar{f}_2(\tilde{X}_0^h) \bar{f}_1(\tilde{X}_k^h)].$$

The proof is thus completed. \square

We prove that the convergence results in Theorem 1 hold for $\bar{b}_{h,N}^{\pm}$. The fact that they hold for $\hat{b}_{h,N}^{\pm}$ as well follows from Lemma 3. Let us recall the expressions of Q_{∞}^{\pm} and $Q_{\infty}^{\pm,1}$ in (2.15).

Lemma 5. *For fixed $h > 0$, let $Q_{h,N}^{\pm}$ as in (2.3). Then*

$$\frac{Q_{h,N}^{\pm}}{N} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} hQ_{\infty}^{\pm} = \frac{h|b_{\mp}|}{|b_{+}| + b_{-}} \quad (4.16)$$

and

$$\sqrt{N} \left(\frac{Q_{h,N}^{\pm}}{N} - \frac{h|b_{\mp}|}{|b_{+}| + b_{-}} \right) \xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, h^2 \xi_{\pm}), \quad (4.17)$$

with

$$\xi_{\pm} := \mathbf{E}[(\mathbf{1}_{\{\pm X_0 \geq 0\}} - Q_{\infty}^{\pm})^2] + 2 \sum_{k=1}^{\infty} \mathbf{E}[(\mathbf{1}_{\{\pm X_0 \geq 0\}} - Q_{\infty}^{\pm})(\mathbf{1}_{\{\pm X_k \geq 0\}} - Q_{\infty}^{\pm})]. \quad (4.18)$$

Let now $Q_{h,N}^{\pm,1}$ be the estimator in (2.9). Then

$$\frac{Q_{h,N}^{\pm,1}}{N} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} hQ_{\infty}^{\pm,1} = \frac{\pm h|b_{\mp}| \sigma_{\pm}^2}{2|b_{\pm}|(b_{-} + |b_{+}|)}. \quad (4.19)$$

Proof. Lemma 4-I implies that

$$Q_{h,N}^{\pm}/N \xrightarrow[N \rightarrow \infty]{\text{a.s.}} h\mathbf{E}[\mathbf{1}_{\{\pm X_{\infty} \geq 0\}}],$$

which, with (2.15), proves (4.16). Asymptotic normality (4.17) also holds because of Lemma 4-II. The limit (4.19) follows from Lemma 4-I and (2.15). \square

We are going to prove Lemma 1. We first need an auxiliary result. Let us write

$$\begin{aligned} G(x, h) &:= \mathbf{E}[|X_\infty^h| \mathbf{1}_{\{X_\infty X_\infty^h < 0\}} \mid X_\infty = x] \\ J(x, h) &:= \mathbf{E}[\mathbf{1}_{\{X_\infty X_\infty^h < 0\}} \mid X_\infty = x] = \mathbf{P}(X_\infty X_\infty^h < 0 \mid X_\infty = x). \end{aligned} \quad (4.20)$$

We have the following lemma.

Lemma 6. *The Laplace transform $\mathcal{L}_\lambda G(x, \cdot) = \int_0^\infty e^{-\lambda t} G(x, t) dt$ is*

$$\mathcal{L}_\lambda G(x, \cdot) = \begin{cases} \frac{2\sigma_-^2}{(b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda})^2} \frac{\exp\left(x \frac{-b_- - \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2}\right)}{\frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2} + \frac{b_+ + \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2}} & \text{for } x > 0 \\ \frac{2\sigma_+^2}{(|b_+| + \sqrt{b_+^2 + 2\sigma_+^2 \lambda})^2} \frac{\exp\left(x \frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2}\right)}{\frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2} + \frac{b_+ + \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2}} & \text{for } x < 0 \end{cases}$$

and the Laplace transform $\mathcal{L}_\lambda J(x, \cdot) = \int_0^\infty e^{-\lambda t} J(x, t) dt$ is

$$\mathcal{L}_\lambda J(x, \cdot) = \begin{cases} \frac{2}{b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}} \frac{\exp\left(x \frac{-b_- - \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2}\right)}{\frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2} + \frac{b_+ + \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2}} & \text{for } x > 0 \\ \frac{2}{|b_+| + \sqrt{b_+^2 + 2\sigma_+^2 \lambda}} \frac{\exp\left(x \frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2}\right)}{\frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2} + \frac{b_+ + \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2}} & \text{for } x < 0. \end{cases}$$

Proof. We apply (4.9), with $f(x, y) = \mathbf{1}_{\{xy < 0\}}|y|$. It is possible to exchange the order of integration because $f(x, y) \geq 0$. Using (4.8) we get, for $x < 0$

$$\int_{\mathbb{R}} r(\lambda, x, y) f(x, y) dy = \frac{\psi_\lambda(x)}{W_\lambda} \int_0^\infty m(y) \varphi_\lambda(y) f(x, y) dy. \quad (4.21)$$

The last integral, using (4.2) and (4.5), can be evaluated as

$$\frac{2}{\sigma_+^2} \int_0^\infty \exp\left(y \frac{-|b_+| - \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2}\right) y dy = \frac{2\sigma_+^2}{(|b_+| + \sqrt{b_+^2 + 2\sigma_+^2 \lambda})^2}$$

so that

$$\mathcal{L}_\lambda \mathbf{E}_x[f(x, X_\cdot)] = \frac{2\sigma_+^2}{(|b_+| + \sqrt{b_+^2 + 2\sigma_+^2 \lambda})^2} \frac{\exp\left(x \frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2}\right)}{\frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}}{\sigma_-^2} + \frac{b_+ + \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2}}.$$

For $x > 0$ the computation is analogous. If we now apply (4.9), with $f(x, y) = \mathbf{1}_{\{xy < 0\}}$, which again is positive, using (4.8) we get, for $x < 0$

$$\int_{\mathbb{R}} r(\lambda, x, y) f(x, y) dy = \frac{\psi_\lambda(x)}{W_\lambda} \int_0^\infty m(y) \varphi_\lambda(y) dy. \quad (4.22)$$

The last integral, using (4.2) and (4.5), can be evaluated as

$$\frac{2}{\sigma_+^2} \int_0^\infty \exp\left(y \frac{-|b_+| - \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}{\sigma_+^2}\right) dy = \frac{2}{|b_+| + \sqrt{b_+^2 + 2\sigma_+^2 \lambda}}$$

so that

$$\mathcal{L}_\lambda \mathbf{E}_x[f(x, X)] = \frac{2}{|b_+| + \sqrt{b_+^2 + 2\sigma_+^2\lambda}} \frac{\exp\left(x \frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2\lambda}}{\sigma_-^2}\right)}{\frac{-b_- + \sqrt{b_-^2 + 2\sigma_-^2\lambda}}{\sigma_-^2} + \frac{b_+ + \sqrt{b_+^2 + 2\sigma_+^2\lambda}}{\sigma_+^2}}.$$

For $x > 0$ the computation is analogous. \square

4.3.1 Proof of Lemma 1

Let us start with $L_{h,N}$. Writing $g(x, y) = \mathbf{1}_{\{xy < 0\}}|y|$, from (2.1) we have $L_{h,N} = 2 \sum_{k=0}^{N-1} g(\mathbb{X}_k^h)$. Lemma 4-III implies that

$$L_{h,N}/N \xrightarrow[N \rightarrow \infty]{\text{a.s.}} 2\mathbf{E}[g(\mathbb{X}_\infty^h)],$$

so that, using the tower property of conditional expectation, we get

$$\mathbf{E}[g(\mathbb{X}_\infty^h)] = \mathbf{E}[|X_\infty^h| \mathbf{1}_{\{X_\infty X_\infty^h < 0\}}] = \mathbf{E}[G(X_\infty, h)]$$

with G in (4.20). The proof strategy consists in computing this quantity via Laplace transform. The Laplace transform $\mathcal{L}_\lambda G(x, \cdot) = \int_0^\infty e^{-\lambda t} G(x, t) dt$ is computed in Lemma 6. Using the stationary distribution (4.2), explicit computations give

$$\begin{aligned} \mathbf{E}[\mathcal{L}_\lambda G(X_\infty, \cdot)] &= \frac{1}{M(\mathbb{R})} \int_{-\infty}^0 \mathcal{L}_\lambda G(x, \cdot) m(x) dx + \frac{1}{M(\mathbb{R})} \int_0^\infty \mathcal{L}_\lambda G(x, \cdot) m(x) dx \\ &= \frac{|b_+| |b_-|}{b_- + |b_+|} \frac{1}{\lambda^2}. \end{aligned}$$

By Tonelli's theorem, it holds that

$$\mathbf{E}[\mathcal{L}_\lambda G(X_\infty, \cdot)] = \int_0^\infty e^{-\lambda t} \mathbf{E}[G(X_\infty, t)] dt = \mathcal{L}_\lambda \mathbf{E}[G(X_\infty, \cdot)].$$

Therefore, inverting the Laplace transform, we obtain

$$\mathbf{E}[g(\mathbb{X}_\infty^h)] = \mathbf{E}[G(X_\infty, h)] = \mathcal{L}_h^{-1} \left(\frac{|b_+| |b_-|}{b_- + |b_+|} \frac{1}{(\cdot)^2} \right) = \frac{|b_+| |b_-|}{b_- + |b_+|} h, \quad (4.23)$$

which proves (2.12).

Let us now consider $\hat{L}_{h,N}$. One can rewrite $\hat{L}_{h,N}$ as

$$\hat{L}_{h,N} = -\frac{3\sqrt{\pi} \sigma_- + \sigma_+}{2\sqrt{2}} \frac{1}{\sigma_- \sigma_+} \frac{1}{\sqrt{h}} \sum_{k=0}^{N-1} X_{kh} X_{(k+1)h} \mathbf{1}_{\{X_{kh} X_{(k+1)h} < 0\}}$$

so that Lemma 4-III implies that as $N \rightarrow \infty$, $\hat{L}_{h,N}/N$ converges \mathbf{P} -a.s. to

$$-\frac{3\sqrt{\pi} \sigma_- + \sigma_+}{2\sqrt{2}} \frac{1}{\sigma_- \sigma_+} \frac{1}{\sqrt{h}} \mathbf{E}[|X_\infty| g(\mathbb{X}_\infty^h)] = -\frac{3\sqrt{\pi} \sigma_- + \sigma_+}{2\sqrt{2}} \frac{1}{\sigma_- \sigma_+} \mathbf{E}[X_\infty X_\infty^h \mathbf{1}_{\{X_\infty X_\infty^h < 0\}}].$$

Using Lemma 6 and the stationary distribution (4.2), explicit computations give

$$\begin{aligned} \mathbf{E}[\mathcal{L}_\lambda(|X_\infty| G(X_\infty, \cdot))] &= \frac{1}{M(\mathbb{R})} \int_{-\infty}^0 \mathcal{L}_\lambda G(x, \cdot) m(x) |x| dx + \frac{1}{M(\mathbb{R})} \int_0^\infty \mathcal{L}_\lambda G(x, \cdot) m(x) |x| dx \\ &= \frac{|b_+| |b_-|}{|b_+| + b_-} \frac{1}{\lambda^3} \frac{(-b_- + \sqrt{b_-^2 + 2\sigma_-^2\lambda})(b_+ + \sqrt{b_+^2 + 2\sigma_+^2\lambda})}{b_- + \sqrt{b_-^2 + 2\sigma_-^2\lambda} + |b_+| + \sqrt{b_+^2 + 2\sigma_+^2\lambda}} \\ &= \sqrt{2} \frac{|b_+| |b_-|}{|b_+| + b_-} \frac{\sigma_- \sigma_+}{\sigma_- + \sigma_+} \frac{1}{\lambda^{5/2}} + \frac{|b_+| |b_-|}{|b_+| + b_-} \left(\frac{\sigma_- b_+ - \sigma_+ b_-}{\sigma_- + \sigma_+} - \frac{(b_- + |b_+|) \sigma_- \sigma_+}{(\sigma_- + \sigma_+)^2} \right) \frac{1}{\lambda^3} + O(\lambda^{-7/2}) \end{aligned}$$

where with $O(\lambda^{-7/2})$ we denote a function $f(\lambda)$ s.t. $\lambda^{7/2}f(\lambda) \rightarrow C \in \mathbb{R}$, as $\lambda \rightarrow \infty$. Therefore, inverting the Laplace transform, using linearity and a Tauberian theorem for $O(\lambda^{-7/2})$ we obtain

$$\begin{aligned}
\mathbf{E}[|X_\infty|g(\mathbb{X}_\infty^h)] &= \mathcal{L}_h^{-1}\mathbf{E}[\mathcal{L}_\lambda(|X_\infty|G(X_\infty, \cdot))] \\
&= \frac{\sqrt{2}}{\Gamma(5/2)} \frac{|b_+|b_-}{|b_+| + b_-} \frac{\sigma_- \sigma_+}{\sigma_- + \sigma_+} h^{3/2} \\
&+ \frac{1}{\Gamma(3)} \frac{|b_+|b_-}{|b_+| + b_-} \left(\frac{\sigma_- b_+ - \sigma_+ b_-}{\sigma_- + \sigma_+} - \frac{(b_- + |b_+|)\sigma_- \sigma_+}{(\sigma_- + \sigma_+)^2} \right) h^2 + O(h^{5/2}) \\
&=: \hat{L}(h) \left(\frac{2\sqrt{2}}{3\sqrt{\pi}} \frac{\sigma_- \sigma_+}{\sigma_- + \sigma_+} h^{3/2} \right)
\end{aligned} \tag{4.24}$$

as $h \rightarrow 0$, where Γ denotes the Γ -function. This implies (2.14).

Let us now consider $\bar{L}_{h,N}$. Let $j(x, y) := \mathbf{1}_{\{xy < 0\}}$, then $\bar{L}_{h,N} = \sqrt{\frac{\pi}{2}} \frac{\sigma_+ + \sigma_-}{2} \sqrt{h} \sum_{k=0}^{N-1} j(\mathbb{X}_k^h)$. The proof works the same as in the case of $L_{h,N}$ but g, G are now j, J . Lemma 6, the stationary distribution (4.2) and explicit computations give

$$\begin{aligned}
\mathbf{E}[\mathcal{L}_\lambda J(X_\infty, \cdot)] &= \frac{1}{M(\mathbb{R})} \int_{-\infty}^0 \mathcal{L}_\lambda J(x, \cdot) m(x) dx + \frac{1}{M(\mathbb{R})} \int_0^\infty \mathcal{L}_\lambda J(x, \cdot) m(x) dx \\
&= \frac{8|b_+|b_-}{b_- + |b_+|} \frac{1}{2\lambda} \frac{1}{|b_+| + \sqrt{b_+^2 + 2\sigma_+^2 \lambda} + b_- + \sqrt{b_-^2 + 2\sigma_-^2 \lambda}} \\
&= \sqrt{2} \frac{2|b_+|b_-}{b_- + |b_+|} \frac{1}{\sigma_+ + \sigma_-} \frac{1}{\lambda^{3/2}} - \frac{2|b_+|b_-}{(\sigma_+ + \sigma_-)^2} \frac{1}{\lambda^2} + O(\lambda^{-5/2})
\end{aligned}$$

as $\lambda \rightarrow \infty$. Therefore, inverting the Laplace transform and using a Tauberian theorem like before we obtain

$$\begin{aligned}
\mathbf{E}[j(\mathbb{X}_\infty^h)] &= \mathcal{L}_h^{-1}\mathbf{E}[\mathcal{L}_\lambda J(X_\infty, \cdot)] \\
&= \frac{1}{\Gamma(3/2)} \frac{|b_+|b_-}{b_- + |b_+|} \frac{2\sqrt{2}}{\sigma_+ + \sigma_-} \sqrt{h} + \frac{1}{\Gamma(2)} \frac{2|b_+|b_-}{(\sigma_+ + \sigma_-)^2} h + O(h^{3/2}) \\
&=: \bar{L}(h) \left(\frac{\sqrt{2}}{\sqrt{\pi}} \frac{2}{\sigma_+ + \sigma_-} h^{1/2} \right)
\end{aligned} \tag{4.25}$$

which implies (2.13).

4.3.2 Proof of Theorem 1 and Theorem 4

Lemma 7. *It holds*

$$\frac{1}{\sqrt{N}} \left(\frac{1}{hQ_\infty^+} (-L_{h,N}/2 - Q_{h,N}^+ b_+) \right) \xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, \Gamma),$$

where

$$\Gamma = \mathbf{E}[q(\tilde{\mathbb{X}}_0^h)q^T(\tilde{\mathbb{X}}_0^h)] + \sum_{k=1}^{\infty} \mathbf{E}[q(\tilde{\mathbb{X}}_0^h)q^T(\tilde{\mathbb{X}}_k^h)] + \mathbf{E}[q(\tilde{\mathbb{X}}_k^h)q^T(\tilde{\mathbb{X}}_0^h)] \tag{4.26}$$

with

$$q(x, y) = \begin{pmatrix} q_1(x, y) \\ q_2(x, y) \end{pmatrix} = \begin{pmatrix} \frac{b_- + |b_+|}{b_-} (-\mathbf{1}_{\{xy < 0\}}|y| - b_+ h \mathbf{1}_{\{x > 0\}}) \\ \frac{b_- + |b_+|}{b_+} (\mathbf{1}_{\{xy < 0\}}|y| - b_- h \mathbf{1}_{\{x < 0\}}) \end{pmatrix}. \tag{4.27}$$

Proof. Note that

$$\left(\frac{\frac{1}{hQ_\infty^+} (-L_{h,N}/2 - Q_{h,N}^+ b_+)}{\frac{1}{hQ_\infty^-} (L_{h,N}/2 - Q_{h,N}^- b_-)} \right) = \sum_{k=0}^{N-1} q(\mathbb{X}_k^h).$$

From (4.25) and (2.15) we have $\mathbf{E}[q(\mathbb{X}_\infty^h)] = 0$. So, Lemma 4-III gives

$$\frac{1}{N} \sum_{k=0}^{N-1} q(\mathbb{X}_k^h) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} 0,$$

and Lemma 4-IV, using again $\mathbf{E}[q(\mathbb{X}_\infty^h)] = 0$, gives the convergence in the statement. \square

Proof of Theorem 1. Consistency is clear from (4.16) and (2.12). The limit error of estimator $(\bar{b}_{h,N}^-, \bar{b}_{h,N}^+)$ can be rewritten with some manipulations as

$$\sqrt{N} \begin{pmatrix} \bar{b}_{h,N}^+ - b_+ \\ \bar{b}_{h,N}^- - b_- \end{pmatrix} = \frac{1}{\sqrt{N}} \begin{pmatrix} \frac{1}{hQ_\infty^+} \left(-\frac{L_{h,N}}{2} - Q_{h,N}^+ b_+ \right) \\ \frac{1}{hQ_\infty^-} \left(\frac{L_{h,N}}{2} - Q_{h,N}^- b_- \right) \end{pmatrix} + \begin{pmatrix} \frac{\sqrt{N}}{2hQ_\infty^+} \left(hQ_\infty^+ - \frac{Q_{h,N}^+}{N} \right) \left(-\frac{L_{h,N}}{Q_{h,N}^+} + \frac{L_\infty}{Q_\infty^+} \right) \\ \frac{\sqrt{N}}{2hQ_\infty^-} \left(hQ_\infty^- - \frac{Q_{h,N}^-}{N} \right) \left(\frac{L_{h,N}}{Q_{h,N}^-} - \frac{L_\infty}{Q_\infty^-} \right) \end{pmatrix}.$$

The last summand vanishes because of (4.16), (4.17), and (2.12). The convergence of the first summand follows from Lemma 7 and implies Theorem 1-ii) (also using (2.15)). \square

Proof of Theorem 4. Consistency follows from (4.16), (4.19) and (2.12). To prove asymptotic normality of the positive side we write

$$\begin{aligned} \sqrt{N}((\bar{\sigma}_{h,N}^+)^2 - \sigma_+^2) &= \frac{\sqrt{N}}{2} \left(\frac{L_{h,N} Q_{h,N}^{+,1}}{(Q_{h,N}^+)^2} - \frac{L_\infty Q_\infty^{+,1}}{(Q_\infty^+)^2} \right) \\ &= \frac{\sqrt{N}}{2} \frac{L_{h,N}/N Q_{h,N}^{+,1}/N (hQ_\infty^+)^2 - hL_\infty hQ_\infty^{+,1} (Q_{h,N}^+/N)^2}{(Q_{h,N}^+/N hQ_\infty^+)^2} \\ &= \frac{\sqrt{N}}{2} \frac{(L_{h,N}/N - hL_\infty) Q_\infty^{+,1} (Q_\infty^+)^2 + L_\infty (Q_{h,N}^{+,1}/N - hQ_\infty^{+,1}) (Q_\infty^+)^2}{h(Q_\infty^+)^4} + \\ &\quad + \frac{\sqrt{N}}{2} \frac{2L_\infty Q_\infty^{+,1} Q_\infty^+ (hQ_\infty^+ - Q_{h,N}^+/N)}{h(Q_\infty^+)^4} + o_N(1) \\ &= \frac{\sqrt{N}}{2} \left(\frac{L_{h,N}}{N} \frac{\sigma_+^2 (b_- + |b_+|)}{2h|b_+|b_-} + \frac{Q_{h,N}^{+,1}}{N} \frac{2|b_+|(b_- + |b_+|)}{hb_-} - \frac{Q_\infty^{+,1}}{N} \frac{2\sigma_+^2 (b_- + |b_+|)}{hb_-} \right) + o_N(1) \end{aligned}$$

where with $o_N(1)$ we mean a quantity that goes to 0 in probability as $N \rightarrow \infty$, that may vary from line to line. A similar computation also holds for the negative side, so that we can write

$$\sqrt{N} \begin{pmatrix} \bar{\sigma}_{h,N}^+ - \sigma_+^2 \\ \bar{\sigma}_{h,N}^- - \sigma_-^2 \end{pmatrix} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} v(\mathbb{X}_k^h) + o_N(1)$$

with

$$v(x, y) = \frac{(b_- + |b_+|)}{b_- |b_+|} \left(\mathbf{1}_{\{xy < 0\}} \frac{|y|}{2h} \begin{pmatrix} \sigma_+^2 \\ \sigma_-^2 \end{pmatrix} + |x| \begin{pmatrix} b_+^2 \mathbf{1}_{\{x > 0\}} \\ b_-^2 \mathbf{1}_{\{x < 0\}} \end{pmatrix} - \begin{pmatrix} \sigma_+^2 |b_+| \mathbf{1}_{\{x > 0\}} \\ \sigma_-^2 b_- \mathbf{1}_{\{x < 0\}} \end{pmatrix} \right).$$

It can be verified from (4.25) and (2.15) that $\mathbf{E}[v(\mathbb{X}_\infty^h)] = 0$. Using Lemma 4-IV allows to conclude, analogously to the proof of Lemma 7, this time with asymptotic covariance

$$\Gamma_\sigma = \mathbf{E}[v(\tilde{\mathbb{X}}_0^h) v^T(\tilde{\mathbb{X}}_0^h)] + \sum_{k=1}^{\infty} \mathbf{E}[v(\tilde{\mathbb{X}}_0^h) v^T(\tilde{\mathbb{X}}_k^h)] + \mathbf{E}[v(\tilde{\mathbb{X}}_k^h) v^T(\tilde{\mathbb{X}}_0^h)], \quad (4.28)$$

so that the theorem is proved. \square

4.4 Proof of Theorem 2: maximum likelihood estimator, high frequency and long time horizon

In Lejay and Pigato [2020] it was proved that the MLE from continuous time observation on the time interval $[0, T]$, $T \in (0, \infty)$ is given by

$$\beta_T := (\beta_T^+, \beta_T^-) = \left(\frac{M_T^+}{Q_T^+}, \frac{M_T^-}{Q_T^-} \right). \quad (4.29)$$

Note that it coincides with the estimator obtained by maximizing the quasi-likelihood function, which is defined as the likelihood, but with diffusion coefficient set to 1. The discrete version of the estimator has been provided as well and is (2.5) as the quantity maximizing the *discretized likelihood*:

$$G_{h,N}(b_+, b_-) = \exp \left(\sum_{i=0}^{N-1} \frac{b(X_{ih})}{(\sigma(X_{ih}))^2} (X_{(i+1)h} - X_{ih}) - \frac{h}{2} \sum_{i=0}^{N-1} \frac{(b(X_{ih}))^2}{(\sigma(X_{ih}))^2} \right).$$

We now state the results about the asymptotic behavior in long time of the estimators based on continuous observations.

Lemma 8 (cf. Proposition 3 in Lejay and Pigato [2020]). *Assume the process is ergodic. Let $\pm \in \{+, -\}$. The MLE (4.29)*

i) *is consistent: $\beta_T^\pm - b_\pm \xrightarrow[T \rightarrow \infty]{\text{a.s.}} 0$, and*

ii) *satisfies the following CLT: $\sqrt{T}(\beta_T^\pm - b_\pm) \xrightarrow[T \rightarrow \infty]{\text{stably}} N^\pm$ where N^+, N^- are two mutually independent, independent of X , Gaussian random variables with variance respectively $\sigma_\pm^2 \frac{|b_+| + |b_-|}{|b_\pm|}$.*

The proof of Theorem 2 exploits the latter lemma and the key Lemma 9 which is based on a bound for the transition density proposed in Lemma 10. More precisely for all $N \in \mathbb{N}$ it holds

$$\widehat{b}_{h_N, N}^\pm - b_\pm = \widehat{b}_{h_N, N}^\pm - \beta_{T_N}^\pm + \beta_{T_N}^\pm - b_\pm.$$

We then handle the second term of the sum with Lemma 8. Equations (4.29) and (2.5) imply for $j \leq 1$ that

$$\begin{aligned} \sqrt{T_N^j} \left(\widehat{b}_{h_N, N}^\pm - \beta_{T_N}^\pm \right) &= \frac{M_{h_N, N}^\pm - M_{T_N}^\pm}{\sqrt{T_N}} \frac{\sqrt{T_N^{1+j}}}{Q_{h_N, N}^\pm} + \frac{M_{T_N}^\pm}{\sqrt{T_N}} \sqrt{T_N^{1+j}} \left(\frac{1}{Q_{h_N, N}^\pm} - \frac{1}{Q_{T_N}^\pm} \right) \\ &= \frac{M_{h_N, N}^\pm - M_{T_N}^\pm}{\sqrt{T_N}} \frac{\sqrt{T_N^{1+j}}}{Q_{h_N, N}^\pm} + \frac{M_{T_N}^\pm}{\sqrt{T_N}} \frac{Q_{T_N}^\pm - Q_{h_N, N}^\pm}{\sqrt{T_N}} \frac{T_N^2}{Q_{h_N, N}^\pm Q_{T_N}^\pm}. \end{aligned} \quad (4.30)$$

The latter converges in probability to 0 by Lemma 9 and the fact that $Q_{T_N}^\pm/T_N$ converges in probability (ergodicity) and $M_{T_N}^\pm/\sqrt{T_N}$ converges in law (ergodicity and martingale CLT in Crimaldi and Pratelli [2005]). Lemma 9 is similar to Lemma 4 in Mazzonetto and Pigato [2024]. Nevertheless, here we allow a deterministic initial condition using the density bound in Lemma 10.

Lemma 9. *Let $\alpha \in [0, 1]$ be fixed and let X be the solution to (1.1), with X_0 deterministic, and let $(T_N)_{N \in \mathbb{N}}, (h_N)_{N \in \mathbb{N}}$ be positive sequences satisfying $h_N = T_N/N$ such that, as $N \rightarrow \infty$, that $T_N \rightarrow \infty$ and $T_N^{2(1-\alpha)} h_N \rightarrow 0$. Then*

$$\limsup_{N \rightarrow \infty} T_N^{-\alpha} \mathbf{E} [|Q_{T_N}^\pm - Q_{h_N, N}^\pm|] = 0 \quad \text{and} \quad \limsup_{N \rightarrow \infty} T_N^{-\alpha} \mathbf{E} [|M_{T_N}^\pm - M_{h_N, N}^\pm|] = 0$$

where $Q_{T_N}^\pm, Q_{h_N, N}^\pm, M_{T_N}^\pm, M_{h_N, N}^\pm$ are defined in (2.2) and (2.3).

Proof of Lemma 9. In this proof we use the round ground notation $[t]_N := t_k$ if $t \in [t_k, t_{k+1})$, where $t_k := kh_N$. Moreover, without loss of generality, we assume $1 \leq T_N \leq N$ for all $N \in \mathbb{N}$.

Let us first note for every $t \in (0, \infty)$ that

$$\mathbf{1}_{\{\pm X_t > 0\}} - \mathbf{1}_{\{\pm X_{[t]_N} > 0\}} = \mp \operatorname{sgn}(X_{[t]_N}) \mathbf{1}_{\{X_{[t]_N} X_t < 0\}}.$$

Hence

$$\mathbf{E} [|Q_{T_N}^\pm - Q_{h_N, N}^\pm|] \leq \int_0^{T_N} \mathbf{E} [\mathbf{1}_{\{X_{[t]_N} X_t < 0\}}] dt = \int_0^{T_N} \mathbf{P}(X_{[t]_N} X_t < 0) dt. \quad (4.31)$$

Analogously, triangular inequality, Hölder's inequality, and Itô-isometry imply that

$$\begin{aligned} & \mathbf{E} [|M_{T_N}^\pm - M_{h_N, N}^\pm|] \\ & \leq (|b_-| \vee |b_+|) \int_0^{T_N} \mathbf{E} [\mathbf{1}_{\{X_t X_{[t]_N} < 0\}}] dt + (\sigma_- \vee \sigma_+) \left(\int_0^{T_N} \mathbf{E} [\mathbf{1}_{\{X_t X_{[t]_N} < 0\}}] dt \right)^{1/2}. \end{aligned} \quad (4.32)$$

Hence, the proof of Lemma 9, which consists in showing that (4.31) and (4.32) are $o(T_N^\alpha)$, reduces to prove

$$\int_0^{T_N} \mathbf{P}(X_{[t]_N} X_t < 0) dt \text{ is } o(T_N^\alpha). \quad (4.33)$$

Step 1: Let $s, t \in (0, \infty)$ be fixed such that $s \leq t$. We show that there exist a constant $C \in (0, \infty)$ depending only on b_\pm, σ_\pm such that for N large enough

$$\mathbf{E} [\mathbf{1}_{\{X_s X_t < 0\}}] \leq C \left(1 + \frac{1}{\sqrt{s}} \right) \sqrt{t - s}. \quad (4.34)$$

In order to prove this, let us fix $\pm \in \{-, +\}$. Note that $\mathbf{E} [\mathbf{E} [\mathbf{1}_{\{\pm X_t < 0\}} \mathbf{1}_{\{\pm X_s > 0\}} | X_s]]$ is bounded by $\mathbf{E} [\mathbf{P}(\tau_s^\pm \leq t - s)]$ where τ_s^\pm is the first hitting time of the level 0 of the process solution to the Brownian motion with drift

$$\xi_u = X_s + b_\pm u + \sigma_\pm W_u^s = X_s + b(X_s)u + \sigma(X_s)W_u^s$$

with W^s a Brownian motion independent of $\sigma(X_v, v \in [0, s])$. Then

$$\mathbf{E} [\mathbf{1}_{\{\pm X_t < 0\}} \mathbf{1}_{\{\pm X_s > 0\}} | X_s] \leq \mathbf{P}(\tau_s^\pm \leq t - s) = \int_0^{t-s} \frac{|X_s|}{\sigma_\pm \sqrt{2\pi u^3}} \exp\left(-\frac{(X_s - b_\pm u)^2}{2\sigma_\pm^2 u}\right) du.$$

Let $f_\pm(y; u) := \exp\left(-\frac{y^2}{2\sigma_\pm^2 u}\right)$. We apply Lemma 10 and obtain for some non-negative constant C :

$$\begin{aligned} & \mathbf{E} [\mathbf{1}_{\{\pm X_t < 0\}} \mathbf{1}_{\{\pm X_s > 0\}}] \leq \int_{\mathbb{R}_\pm} q_{(b_-, b_+)}(s, X_0, x) \int_0^{t-s} \frac{|x|}{\sigma_\pm \sqrt{2\pi u^3}} f_\pm(x - b_\pm u; u) du dx \\ & \leq C \frac{1 + \sqrt{s}}{\sqrt{s}} \int_0^{t-s} \int_{\mathbb{R}_\pm} \frac{|x + 2b_\pm u|}{\sigma_\pm^2 \sqrt{2\pi u^3}} f_\pm(x; u) dx du \\ & \leq C \frac{1 + \sqrt{s}}{\sqrt{s}} \int_0^{t-s} \int_{\mathbb{R}_\pm} \frac{|x|}{\sqrt{2\pi u^3}} e^{-\frac{x^2}{2u}} dx du + \frac{2|b_\pm|}{\sigma_\pm^2} C \frac{1 + \sqrt{s}}{\sqrt{s}} \int_0^{t-s} \int_{\mathbb{R}_\pm} \frac{1}{\sqrt{2\pi u}} e^{-\frac{x^2}{2u}} dx du \\ & = C \frac{1 + \sqrt{s}}{\sqrt{s}} \left(\frac{1}{\sqrt{2\pi}} \int_0^{t-s} \frac{1}{\sqrt{u}} du + \frac{2|b_\pm|}{\sigma_\pm^2} (t - s) \right) \\ & = \left(\sqrt{\frac{2}{\pi}} + \frac{2|b_\pm|}{\sigma_\pm^2} \sqrt{t - s} \right) C \left(1 + \frac{1}{\sqrt{s}} \right) \sqrt{t - s}. \end{aligned}$$

This completes the proof of the first step.

Step 2: Proof of (4.33).

For all $t \in [t_1, \infty)$ the tower property and (4.34) imply, in the notation of (4.34), that

$$\mathbf{P}(X_{\lfloor t \rfloor_N} X_t < 0) = \mathbf{E} \left[\mathbf{E} \left[\mathbf{1}_{\{X_{\lfloor t \rfloor_N} X_t < 0\}} \mid X_{\lfloor t \rfloor_N} \right] \right] \leq C \left(1 + \frac{1}{\sqrt{\lfloor t \rfloor_N}} \right) \sqrt{t - \lfloor t \rfloor_N}.$$

Then,

$$\frac{1}{T_N^\alpha} \int_0^{T_N} \mathbf{P}(X_{\lfloor t \rfloor_N} X_t < 0) dt \leq \frac{2h_N}{T_N^\alpha} + C \frac{1}{T_N^\alpha} \int_{2h_N}^{T_N} \sqrt{t - \lfloor t \rfloor_N} dt + C \frac{1}{T_N^\alpha} \int_{2h_N}^{T_N} \frac{\sqrt{t - \lfloor t \rfloor_N}}{\sqrt{\lfloor t \rfloor_N}} dt.$$

On the right-hand-side of the latter inequality, the first term vanishes as $N \rightarrow \infty$, the second term is bounded by

$$C \frac{1}{T_N^\alpha} \int_0^{T_N} \sqrt{t - \lfloor t \rfloor_N} dt \leq C \sqrt{h_N T_N^{2(1-\alpha)}},$$

and the last term is bounded by

$$C \frac{1}{T_N^\alpha} \int_{2h_N}^{T_N} \frac{\sqrt{t - \lfloor t \rfloor_N}}{\sqrt{t - h_N}} dt \leq C \frac{\sqrt{h_N}}{T_N^\alpha} \int_0^{T_N} \frac{1}{\sqrt{u}} du = C \sqrt{h_N T_N^{1-2\alpha}}.$$

The proof is thus completed. \square

Let $q_{(b_+, b_-)}(t, x, y)$ denote the transition density of the drifted oscillating Brownian motion (1.1) with diffusion coefficient σ in (1). To bound the density, we take inspiration from the article Downes [2009].

Note that the transition density of the oscillating Brownian motion without drift, $q_{(0,0)}(t, x, y)$, is known. It was provided in Keilson and Wellner [1978], and it satisfies for instance that $q_{(0,0)}(t, x, y) \leq Ct^{-1/2}$ for some constant $C \in (0, \infty)$ depending on the parameters.

Lemma 10. *There exist $c, C, K \in (0, \infty)$ depending on the parameters b_\pm, σ_\pm such that for all $y \in \mathbb{R}, t \in (0, \infty)$ it holds that*

$$q_{(b_+, b_-)}(t, x, y) \leq C e^{yb(y)/\sigma^2(y) - xb(x)/\sigma^2(x)} (1 + \sqrt{t}) q_{(0,0)}(t, x, y).$$

Moreover, $q_{(0,0)}(t, x, y) \leq \frac{1}{\sigma(y)} \left(1 + \frac{|\sigma_- - \sigma_+|}{\sigma_- + \sigma_+} \right) \frac{1}{\sqrt{2\pi t}} e^{-\left(\frac{y}{\sigma(y)^2} - \frac{x}{\sigma(x)^2}\right)/2t}$.

Proof of Lemma 10. Note that, if Y is a drifted oscillating Brownian motion (1.1) with diffusion coefficient σ in (1) and drift $b(x)$, then $X/\sigma(X)$ is a skew Brownian motion (SBM) with drift $b(x)/\sigma(x)$ and skewness parameter $\frac{\sigma_- - \sigma_+}{\sigma_- + \sigma_+}$. This follows from Itô-Tanaka formula, for a proof see e.g. Mazzone [2019]. Moreover, if $p_{(b_+, b_-)}(t, x, y)$ denotes the transition density of the $\beta := \frac{\sigma_- - \sigma_+}{\sigma_- + \sigma_+}$ -SBM with drift b , then

$$q_{(b_+, b_-)}(t, x, y) = p_{(b_+/\sigma_+, b_-/\sigma_-)}(t, x/\sigma(x), y/\sigma(y))/\sigma(y).$$

In this proof we only consider X a skew-BM with piecewise constant drift b (abuse of notation because b is actually b/σ) and skewness parameter β such that $X_0 = x_0$. By Girsanov's theorem there exists a measure \mathbf{Q} absolutely continuous with respect to \mathbf{P} such that X under \mathbf{Q} is a driftless β -SBM with $X_0 = x_0$ and for all f measurable and bounded it holds that

$$\mathbf{E}[f(X_t)] = \mathbf{E}_{\mathbf{Q}} \left[\exp \left(\int_0^t b(X_s) dW_s - \frac{1}{2} \int_0^t (b(X_s))^2 ds \right) f(X_t) \right].$$

Since $b' \equiv 0$, Itô-Tanaka formula applied to the function $B(y) := \int_0^y b(x) dx$ and the driftless β -SBM X_t implies that \mathbf{P} -a.s.

$$\int_0^t b(X_s) dW_s - \frac{1}{2} \int_0^t (b(X_s))^2 ds = B(X_t) - B(X_0) - \frac{1}{2} \int_0^t (b(X_s))^2 + b'(X_s) ds - C_{b,\sigma} L_t^0(X)$$

where $C_{b,\sigma} := \frac{b_+\sigma_- - b_-\sigma_+}{\sigma_- + \sigma_+} \in (-\infty, 0)$. Therefore,

$$\mathbf{E}[f(X_t)] = \mathbf{E}_{\mathbf{Q}}[\exp(B(X_t) - B(X_0) - C_{b,\sigma} L_t^0(X)) f(X_t)].$$

Then

$$\mathbf{E}[f(X_t)] \leq e^{-B(X_0)} \mathbf{E}_{\mathbf{Q}}[e^{B(X_t)} e^{-C_{b,\sigma} L_t^0(X)} f(X_t)] \leq e^{-B(X_0)} \mathbf{E}_{\mathbf{Q}}[e^{B(X_t)} f(X_t) \mathbf{E}_{\mathbf{Q}}[e^{-C_{b,\sigma} L_t^0(X)} | X_t]].$$

To complete the proof we need to show that

$$\mathbf{E}_{\mathbf{Q}}[e^{-C_{b,\sigma} L_t^0(X)} | X_t] \leq 1 + \sqrt{2\pi t} |C_{b,\sigma}|. \quad (4.35)$$

Let us first assume (4.35). Then, for every f measurable and bounded

$$\mathbf{E}[f(X_t)] \leq e^{-B(X_0)} \left(1 + \sqrt{2\pi t} \frac{|b_+\sigma_- - b_-\sigma_+|}{\sigma_- + \sigma_+}\right) \mathbf{E}_{\mathbf{Q}}[f(X_t)]$$

and we complete the proof as in [Downes, 2009, Corollary 5]: to consider the transition density we take derivative of the partition function of X_t . More precisely, we consider $f = \mathbf{1}_{(y-\varepsilon, y+\varepsilon]}$ for $\varepsilon > 0$, divide for ε and take the limit on both sides as $\varepsilon \rightarrow 0$. Thus,

$$p_{(b_+, b_-)}(t, X_0, y) \leq e^{B(y) - B(X_0)} \left(1 + \sqrt{2\pi t} \frac{|b_+\sigma_- - b_-\sigma_+|}{\sigma_- + \sigma_+}\right) p_{(0,0)}(t, X_0, y)$$

and the transition density of a β -SBM, is known (see Walsh [1978]):

$$p_{(0,0)}(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right) + \beta \operatorname{sgn}(y) \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(|x| + |y|)^2}{2t}\right) \quad (4.36)$$

which is bounded from above by $\frac{1+|\beta|}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right)$.

To conclude, we now prove (4.35). In what follows let $C := |C_{b,\sigma}|$.

Let $\rho(t, x, y)$ be the joint transition density of the skew BM $Y_t = x + W_t + \frac{\sigma_- - \sigma_+}{\sigma_- + \sigma_+} L_t^0(Y)$, $t \in [0, \infty)$, and its local time: $(Y_t, L_t^0(Y))$ (see [Étoré and Martinez, 2013, Proposition 1] or Appuhamillage et al. [2011]):

$$\rho_t(\ell, X_0, X_t)(1 - \delta_0(\ell)) = \frac{(1 + \operatorname{sgn}(X_t) \frac{\sigma_- - \sigma_+}{\sigma_- + \sigma_+})(\ell + X_0 + |X_t|)}{\sqrt{2\pi t^3}} e^{-\frac{(\ell + X_0 + |X_t|)^2}{2t}}$$

Recall that

$$\mathbf{E}_{\mathbf{Q}}[e^{cL_t^0(X)} | X_t] = \int_0^\infty e^{c\ell} \frac{\rho_t(\ell, X_0, X_t)}{p_{(0,0)}(t, X_0, X_t)} d\ell$$

where $p_{(0,0)}(t, X_0, X_t)$ is given by (4.36). It holds for all $c \in (0, \infty)$ that

$$\begin{aligned} & e^{\frac{c\ell}{t}} \frac{\rho_t(\ell, X_0, X_t)}{p_{(0,0)}(t, X_0, X_t)} (1 - \delta_0(\ell)) \\ &= \frac{(\ell + X_0 + |X_t| - c) + c}{t} e^{-\frac{(\ell + X_0 + |X_t| - c)^2}{2t}} \frac{(1 + \operatorname{sgn}(X_t) \frac{\sigma_- - \sigma_+}{\sigma_- + \sigma_+}) e^{\frac{(X_0 + |X_t| - c)^2 - 4X_t X_0 \mathbf{1}_{\{X_t > 0\}}}{2t}}}{1 + \operatorname{sgn}(X_t) \frac{\sigma_- - \sigma_+}{\sigma_- + \sigma_+} e^{-2\frac{X_t X_0}{t}} \mathbf{1}_{\{X_0 X_t > 0\}}}. \end{aligned}$$

Note that

$$\int_0^\infty \frac{(\ell + c_2) + c}{t} e^{-\frac{(\ell+c_2)^2}{2t}} d\ell = \int_{c_2}^\infty \frac{\ell + c}{t} e^{-\frac{\ell^2}{2t}} d\ell = e^{-\frac{c_2^2}{2t}} + \frac{c\sqrt{2\pi}}{\sqrt{t}} \Phi\left(-\frac{c_2}{\sqrt{t}}\right)$$

with $c = Ct$ and $c_2 = X_0 + |X_t| - Ct$. Therefore

$$\mathbf{E}_{\mathbf{Q}}\left[e^{CL_t^0(X)}|X_t\right] = e^{-\frac{(X_0+|X_t|-Ct)^2}{2t}} + C\sqrt{t}\sqrt{2\pi}\Phi\left(-\frac{X_0 + |X_t| - Ct}{\sqrt{t}}\right) \leq 1 + \sqrt{2\pi t} C.$$

This completes the proof. □

Acknowledgements. We are grateful to A. Lejay and R. Peyre for discussions. PP acknowledges financial support from University of Rome Tor Vergata via Project AIF - E83C22002610005.

References

- C. Amorino and A. Gloter. Contrast function estimation for the drift parameter of ergodic jump diffusion process. *Scandinavian Journal of Statistics*, 47(2):279–346, 2020.
- A. Ang and A. Timmermann. Regime Changes and Financial Markets. *Annual Review of Financial Economics*, 4:313–337, 2012.
- T. Appuhamillage, V. Bokil, E. Thomann, E. Waymire, and B. Wood. Occupation and local times for skew Brownian motion with applications to dispersion across an interface. *Ann. Appl. Probab.*, 21(1):183–214, 2011.
- R. F. Bass and Z.-Q. Chen. One-dimensional stochastic differential equations with singular and degenerate coefficients. *Sankhyā: The Indian Journal of Statistics*, pages 19–45, 2005.
- M. Ben Alaya and A. Kebaier. Asymptotic Behavior of the Maximum Likelihood Estimator for Ergodic and Nonergodic Square-Root Diffusions. *Stochastic Analysis and Applications*, 31(4):552–573, 2013.
- A. N. Borodin. On the character of convergence to brownian local time. ii. *Probability Theory and Related Fields*, 72:251–277, 1986.
- A. N. Borodin and P. Salminen. *Handbook of Brownian motion-facts and formulae*. Springer Science & Business Media, 2015.
- S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.
- D. Buckner, K. Dowd, and H. Hulley. Arbitrage Problems with Reflected Geometric Brownian Motion. *Finance Stoch*, 28, Jan. 2024.
- C. W. S. Chen, M. K. P. So, and F.-C. Liu. A review of threshold time series models in finance. *Statistics and its Interface*, 4(2):167–181, 2011.
- S. Christensen and C. Strauch. Nonparametric learning for impulse control problems – Exploration vs. exploitation. *The Annals of Applied Probability*, 33(2):1569 – 1587, 2023.
- S. Christensen, C. Strauch, and L. Trottner. Learning to reflect: A unifying approach for data-driven stochastic control strategies. *Bernoulli*, 2023.

- I. Crimaldi and L. Pratelli. Convergence results for multivariate martingales. *Stochastic Process. Appl.*, 115(4):571–577, 2005.
- M. Decamps, M. Goovaerts, and W. Schoutens. Self exciting threshold interest rates models. *Int. J. Theor. Appl. Finance*, 9(7):1093–1122, 2006.
- A. B. Dieker and X. Gao. Positive recurrence of piecewise Ornstein-Uhlenbeck processes and common quadratic Lyapunov functions. *Ann. Appl. Probab.*, 23(4):1291–1317, 08 2013.
- F. Dong and H. Y. Wong. Variance Swaps under the Threshold Ornstein-Uhlenbeck Model. *Appl. Stoch. Model. Bus. Ind.*, 33(5):507–521, Sept. 2017.
- A. N. Downes. Bounds for the transition density of time-homogeneous diffusion processes. *Statistics & probability letters*, 79(6):835–841, 2009.
- P. Étoré and M. Martinez. Exact simulation of one-dimensional stochastic differential equations involving the local time at zero of the unknown process. *Monte Carlo Methods Appl.*, 19(1): 41–71, 2013.
- D. Florens-Zmirou. On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.*, 30(4):790–804, 1993.
- A. Gairat and V. Shcherbakov. Density of skew Brownian motion and its functionals with application in finance. *Mathematical Finance*, 26(4):1069–1088, 2016.
- F. Giorgi, S. Herzel, and P. Pigato. A reinforcement learning algorithm for trading commodities. *Applied Stochastic Models in Business and Industry*, n/a(n/a), 2023.
- S. Hottovy and S. N. Stechmann. Threshold models for rainfall and convection: Deterministic versus stochastic triggers. *SIAM Journal on Applied Mathematics*, 75(2):861–884, 2015.
- Y. Hu and Y. Xi. Parameter estimation for threshold ornstein–uhlenbeck processes from discrete observations. *Journal of Computational and Applied Mathematics*, 411:114264, 2022.
- K. Itô and H. McKean. *Diffusion Processes and their Sample Paths: Reprint of the 1974 Edition*. Classics in Mathematics. Springer Berlin Heidelberg, 1996.
- J. Jacod. Rates of convergence to the local time of a diffusion. *Ann. Inst. H. Poincaré Probab. Statist.*, 34(4):505–544, 1998.
- J. Jacod and P. Protter. *Discretization of processes*, volume 67 of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, 2012.
- J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, second edition, 2003.
- J. Keilson and J. A. Wellner. Oscillating Brownian motion. *J. Appl. Probability*, 15(2):300–310, 1978.
- M. Kessler. Estimation of an Ergodic Diffusion from Discrete Observations. *Scandinavian Journal of Statistics*, 24(2):211–229, 1997.
- Y. A. Kutoyants. On identification of the threshold diffusion processes. *Ann. Inst. Statist. Math.*, 64(2):383–413, 2012.
- J.-F. Le Gall. One-dimensional stochastic differential equations involving the local times of the unknown process. *Stochastic Analysis. Lecture Notes Math.*, 1095:51–82, 1985.

- A. Lejay and P. Pigato. Statistical estimation of the Oscillating Brownian Motion. *Bernoulli*, 24(4B):3568–3602, 2018.
- A. Lejay and P. Pigato. A threshold model for local volatility: evidence of leverage and mean reversion effects on historical data. *International Journal of Theoretical and Applied Finance*, 22(4):1950017, 2019.
- A. Lejay and P. Pigato. Maximum likelihood drift estimation for a threshold diffusion. *Scandinavian Journal of Statistics*, 47(3):609–637, 2020.
- A. Lejay, E. Mordecki, and S. Torres. Two consistent estimators for the skew Brownian motion. *ESAIM Probab. Stat.*, 23:567–583, 2019.
- A. Lipton. Oscillating Bachelier and Black-Scholes Formulas. In *Financial Engineering*. World Scientific, 2018.
- A. Lipton and A. Sepp. Filling the gaps. *Risk Magazine*, pages 66–71, 2011-10.
- S. Mazzonetto. Rates of convergence to the local time of oscillating and skew brownian motions. *arXiv preprint arXiv:1912.04858*, 2019.
- S. Mazzonetto and P. Pigato. Drift estimation of the threshold ornstein-uhlenbeck process from continuous and discrete observations. *Statistica Sinica*, 34:313–336, 2024.
- S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, USA, 2nd edition, 2009. ISBN 0521731828.
- P. P. Mota and M. L. Esquivel. On a continuous time stock price model with regime switching, delay, and threshold. *Quant. Finance*, 14(8):1479–1488, 2014.
- J. Pai and H. Pedersen. Threshold Models of the Term Structure of Interest Rate. In *Joint day Proceedings Volume of the XXXth International ASTIN Colloquium/9th International AFIR Colloquium, Tokyo, Japan*, pages 387–400. 1999.
- P. Pigato. Extreme at-the-money skew in a local volatility model. *Finance and Stochastics*, 23:827–859, 2019.
- N. I. Portenko. The development of I. I. Gikhman’s idea concerning the methods for investigating local behavior of diffusion processes and their weakly convergent sequences. *Teor. Īmov̄r. Mat. Stat.*, (50):6–21, 1994.
- A. Rényi. On stable sequences of events. *Sankhyā Ser. A*, 25:293–302, 1963.
- F. Su and K.-S. Chan. Quasi-likelihood estimation of a threshold diffusion process. *J. Econometrics*, 189(2):473–484, 2015.
- F. Su and K.-S. Chan. Testing for threshold diffusion. *J. Bus. Econom. Statist.*, 35(2):218–227, 2017.
- H. Tong. Threshold models in time series analysis — 30 years on. *Statistics and its Interface*, 4, 2011.
- J. B. Walsh. A diffusion with discontinuous local time. In *Temps locaux*, volume 52-53, pages 37–45. Société Mathématique de France, 1978.
- T.-H. Yu, H. Tsai, and H. Rachinger. Approximate maximum likelihood estimation of a threshold diffusion process. *Computational Statistics & Data Analysis*, 142:106823, 2020.