Parameters estimation of a Threshold CKLS process from continuous and discrete observations

Sara Mazzonetto, Benoît Nieto

Parameters estimation of a Threshold CKLS process from continuous and discrete observations

Sara Mazzonetto* Benoit Nieto†

March 28, 2024

Abstract. We consider a continuous-time process which is self-exciting and ergodic, called threshold Chan–Karolyi–Longstaff–Sanders (CKLS) process. We allow for the presence of several thresholds which determine changes in the dynamics. We study the asymptotic behavior of the maximum and quasi-maximum likelihood estimators of the drift parameters in the case of continuous time and discrete time observations. We show that for high frequency observations and infinite horizon the estimators satisfy the same asymptotic normality property as in the case of continuous time observations. We discuss diffusion coefficient estimation as well. Finally, we apply our estimators to simulated and real data to motivate considering (multiple) thresholds.

Keywords: CIR, CKLS, maximum likelihood, regime-switch, self-exciting process, threshold diffusion.

MSC 2020: primary: 62F12; secondary: 62F03 62M05;

1 Introduction

We deal with parameter estimation of a stochastic process which follows different Chan-Karolyi-Longstaff-Sanders (CKLS) dynamics separately on different intervals (see equation (2.1) for a definition), that we call *threshold CKLS* (T-CKLS) process. T-CKLS is in particular a Self Exciting Threshold model [10]. It excites itself by changing dynamics according to its own position. The class of T-CKLS models includes several threshold and non-threshold models such as Ornstein-Uhlenbeck (OU) or Vasicek, Cox-Ingersoll-Ross (CIR), Black-Scholes, Merton, Constant Elasticity of Variance (CEV) model. Let us recall the equation satisfied by (non-threshold) CKLS:

$$X_t = X_0 + \int_0^t (a - bX_s)\,\mathrm{d}s + \int_0^t \sigma |X_s|^\gamma \,\mathrm{d}B_s \tag{1.1}$$

where $\sigma \in (0, \infty)$, $b \in \mathbb{R}$, $\gamma \in [0, 1]$. Assume also that the deterministic initial condition X_0 and the coefficient a are strictly positive. CKLS was considered for interest rate modeling in [8]. Among statistical study of non-threshold diffusions such as CKLS and CIR let us mention [26] and [4, 5]. The approaches exploit the knowledge of the law of

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

†Ecole Centrale de Lyon, ICJ UMR5208, CNRS, INSA Lyon, Universite Claude Bernard Lyon 1, Université Jean Monnet, 69130 Ecully, France, benoit.nieto@ec-lyon.fr

some functionals of the process. Except for very special threshold diffusions (TDs), this law is not available or not easily exploitable. Therefore, for dealing with TDs, one must consider different techniques which would hold also for non-threshold cases (see Section 5.2 for details).
Recently, several studies have been conducted on the parametric estimation of TDs. Just to mention some results, in the case of continuous time observations [17, 30, 22], high-frequency observations on finite or infinite horizon [20, 24, 25], low frequency ones [33, 12, 25]. TDs attract attention for applications, in financial modelling e.g. [10, 21, 28], population ecology [7], etc. One of the features of the TDs is that they allow for mean reversions even with several mean reversion levels.
In this paper we focus on ergodic T-CKLS. Note that this includes ergodic T-CIR model and allows for different dynamics on fixed intervals: OU on a region, CIR on another, and possibly other special cases of CKLS on another interval. We consider both continuous time observations and discrete observations which are not necessarily equally spaced.
We study both maximum likelihood estimation (MLE) and a quasi maximum likelihood estimation (QMLE) for the drift coefficients and we also propose an estimator for the diffusion coefficients σ based on quadratic variation. We study the asymptotic behavior of these estimators, obtaining a law of large numbers and a central limit theorem for the drift coefficients estimators. We discuss these asymptotic results in long time in the continuous time setting, see Theorem 3.5, and in high frequency and long time in the discrete time setting, see Theorem 4.5. We also study consistency and provide a lower bound for the speed of convergence for the diffusion coefficient estimator in high frequency and long time, see Theorem 4.3.
In the continuous time setting, Theorem 3.5 is new for MLE estimator. The drift QMLE estimator was considered in [30]. In the same paper, numerical studies were conducted on the discretization of the QMLE. The statistical properties of the discretized MLE and QMLE are proven in Theorem 4.5, which is the main result of this paper. Let us mention that the results obtained in this document are new also when the process follows some special dynamics of T-CKLS process, such as T-CIR. In the context of TDs observed in high frequency and long time, Theorem 4.3 on diffusion coefficient estimation is the first result of its kind.
In the case of discrete observations, we assume that the process is in its stationary regime (X_t is distributed according to the stationary distribution for all $t \geq 0$) and it is observed on a time-grid of N observations $0 = t_0 < t_1 < \ldots < t_N = T_N$ with maximal lag between two consecutive observations, say Δ_N, such that $\lim_{N\to\infty} T_N = +\infty$ and $\lim_{N\to\infty} \Delta_N \to 0$. To prove the convergence speed, we require an additional condition of the form $\lim_{N\to\infty} \Delta_N^\alpha T_N = 0$, for some power $\alpha \in (0, 1]$ that depends on the parameter vector γ. This is consistent with the existing literature in the context of parameter estimation of diffusions from discrete observations with and without threshold, e.g. [2, 4, 24] (where $\alpha = 1$). Further comments on our results, their assumptions, possible extensions and comparisons with the literature are provided in Section 5.

Outline. In Section 2, we introduce the model and the quasi-likelihood and likelihood functions associated to the T-CKLS process. In Section 3, we deal with estimation of T-CKLS from continuous observations. We provide some statistical properties related to the drift estimator. The main results of this article are provided in Section 4, which deals with drift and volatility estimation from discrete observations. We study the asymptotic behavior in high frequency and long time for both estimators. In Section 5 we comment

on the results. Numerical experiments are provided in Section 6, where the estimators are implemented and tested on simulated data and US interest rates data. Proofs are collected in Section 7. Further useful results are available in Appendix A.
Throughout the paper, we use the notion of stable convergence, denoted $\xrightarrow{stably}$. Further details on this type of convergence can be found in [14] and [15].

2 The framework: model and assumptions

The T-CKLS process solves the following one-dimensional SDE:

$$X_t = x_0 + \int_0^t (a(X_s) - b(X_s)X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)|X_s|^{\gamma(X_s)}\,\mathrm{d}B_s, \quad t \geq 0, \tag{2.1}$$

with $x_0 > 0$, piecewise constant coefficients a, b, σ and γ possibly discontinuous at levels $0 = r_0 < r_1 < \cdots < r_d < r_{d+1} = +\infty$, $d \in \mathbb{N}$. More precisely, let $I_j := [r_j, r_{j+1})$, for $j \in \{0, \cdots, d\}$, unless $\gamma(0) = 0$, in which case $I_0 = (-\infty, r_1)$. The drift coefficients are given by

$$a(x) = \sum_{j=0}^{d} a_j \mathbb{1}_{I_j}(x) \in \mathbb{R} \quad \text{and} \quad b(x) = \sum_{j=0}^{d} b_j \mathbb{1}_{I_j}(x) \in \mathbb{R}, \tag{2.2}$$

and similarly the volatility coefficients are given by

$$\sigma(x) = \sum_{j=0}^{d} \sigma_j \mathbb{1}_{I_j}(x) > 0 \quad \text{and} \quad \gamma(x) = \sum_{j=0}^{d} \gamma_j \mathbb{1}_{I_j}(x) \geq 0. \tag{2.3}$$

We focus on the case $\gamma(\mathbb{R}) \subseteq [0, 1]$ and we assume that the process spends no time at 0. For existence of a pathwise unique strong solution to (2.1) if $\gamma_0 \in [1/2, 1] \cup 0$, we refer to [19]. In the case $\gamma_0 \in (0, 1/2)$, weak existence and uniqueness in law of solutions which do not spend time at 0 holds for the SDE (2.1) (see e.g. [11]). Pathwise uniqueness and strong existence are more subtle (we refer the reader to [3] for more details). Moreover, T-CKLS is a Markov process (see e.g. [11]).
When $\gamma_0 \in (0, 1]$ the process is always non-negative and 0 is either an unreachable point or a reflecting one (see Lemma 7.1 in Section 7.1). When $\gamma_0 = 0$ the T-CKLS is a threshold OU process, hence one can remove the assumptions that the thresholds r_j are positive and the initial condition x_0 is allowed to be non-positive.
We suppose that $(\gamma_j)_{j=0}^d$ and the thresholds $(r_j)_{j=0}^d$ are known, and we estimate drift and diffusion parameters vectors a, b, σ for continuous time observations and discrete (non necessarily equally spaced) high frequency observations and infinite horizon.

Definition 1 (Ergodicity)**.** *We say that the process is* ergodic*, if it is positive recurrent.*

In this article we consider only the case in which the process is ergodic. For instance, the process is ergodic if we restrict the drift coefficients of the first interval I_0 and the last interval I_d to satisfy

$$(a_0, b_0) \in (0, +\infty) \times \mathbb{R} \quad \text{and} \quad (a_d, b_d) \in \mathbb{R} \times (0, +\infty).$$

More precise restrictions to the parameters are given in Table 4 in Section 7.1.
When the process is ergodic, there exists a stationary distribution (invariant distribution for the transition semigroup), denoted by μ. An expression for the stationary distribution is given in Section 7.1.

Definition 2 (Stationarity)**.** *We say that the process is* stationary *if it is ergodic and X_0 is independent of the driving Brownian motion and it is distributed according to the stationary distribution.*

The maximum likelihood estimation method for the drift parameter also requires parameter restrictions. Assume we have access to an observation of an entire trajectory on the time interval $[0, T]$ of the T-CKLS. We denote by $\theta := (a, b) = (a_j, b_j)_{j=0}^d$ the drift parameters and we assume the thresholds r and the coefficients γ to be known. In the next section we assume σ to be unknown, and we propose an estimator. Yet, in the next lines, the reader should think as if σ is known (replaced by an estimator). We consider two different contrast functions: likelihood and quasi-likelihood. The likelihood function $\theta \mapsto \mathcal{L}_T(\theta; \sigma, \gamma)$ is the Girsanov weight:

$$\mathcal{L}_T(\theta; \sigma, \gamma) = \exp\left(\int_0^T \frac{a(X_s) - b(X_s)X_s}{\sigma(X_s)^2 (X_s)^{2\gamma(X_s)}} \, \mathrm{d}X_s - \frac{1}{2} \int_0^T \frac{(a(X_s) - b(X_s)X_s)^2}{\sigma(X_s)^2 (X_s)^{2\gamma(X_s)}} \, \mathrm{d}s \right). \quad (2.4)$$

Note that it is well defined if the integrals above are well defined. Hence, we further restrict the parameter space $\Theta^{(\mathcal{L})}$ such that $\gamma_0 \in \{0\} \cup (1/2, 1]$, or $\gamma_0 = 1/2$ and $a_0 \geq \sigma_-^2/2$. The likelihood is well defined on $\Theta^{(\mathcal{L})}$ (see Proposition A.1 for details).
We consider the quasi-Likelihood function $\theta \mapsto q\text{-}\mathcal{L}_T(\theta) := \ln \mathcal{L}(\theta; 1, 0)$. Hence,

$$q\text{-}\mathcal{L}_T(\theta) = \int_0^T a(X_s) - b(X_s)X_s \, \mathrm{d}X_s - \frac{1}{2} \int_0^T (a(X_s) - b(X_s)X_s)^2 \, \mathrm{d}s. \quad (2.5)$$

The advantage of this contrast function is the fact that it does not depend on the diffusion's coefficients γ, σ. We denote $\Theta^{(q\text{-}\mathcal{L})}$ the set of parameters such that the quasi-likelihood is well defined. Note that $q\text{-}\mathcal{L}$ is always well defined because the process we consider has continuous trajectories. So, it covers a wider range of parameters than the likelihood function.
Summarizing, in the next sections, we suppose that the parameter $r = (r_j)_{j=1}^d$ and $\gamma = (\gamma_j)_{j=0}^d$ are known. We denote $\theta_\star := (a, b) = (a_j, b_j)_{j=0}^d$ and $\sigma_\star$ the parameters to be estimated. We suppose that the parameters are in $\Theta^{(\mathcal{L})}$, resp. $\Theta^{(q\text{-}\mathcal{L})}$, when dealing with the likelihood, resp. quasi-likelihood function. Moreover, we assume that the process is ergodic (see Table 4 in Section 7.1 for the parameter restrictions ensuring ergodicity).

3 Estimation from continuous time observations

Let $T \in (0, \infty)$, and assume we have at our disposal continuous time observations on the interval $[0, T]$ of a trajectory of the process X solution to the SDE (2.1).
First, we provide estimators which maximize likelihood $\mathcal{L}$ and quasi-likelihood $q\text{-}\mathcal{L}$. Next, we study the asymptotic behavior of the estimators in long time under the assumption that the process is ergodic.

3.1 Estimators expressions

The drift parameters estimators are defined as the maximal argument of the likelihood (3.2) and the quasi-likelihood (3.3):

$$\theta_T^{(\mathcal{L})} := \underset{\theta \in \Theta^{(\mathcal{L})}}{\operatorname{argmax}} \ \ln \mathcal{L}_T(\theta; \sigma, \gamma) \quad \text{and} \quad \theta_T^{(q\text{-}\mathcal{L})} := \underset{\theta \in \Theta^{(q\text{-}\mathcal{L})}}{\operatorname{argmax}} \ q\text{-}\mathcal{L}_T(\theta).$$

We look for expressions for MLE and QMLE in terms of the following quantities:

$$Q_T^{j,m} := \int_0^T X_s^m \mathbb{1}_{I_j}(X_s)\,\mathrm{d}s \quad \text{and} \quad M_T^{j,m} := \int_0^T X_s^m \mathbb{1}_{I_j}(X_s)\,\mathrm{d}X_s \tag{3.1}$$

for $j \in \{0, \cdots, d\}$ and $m \in \{-2\gamma_j, 1-2\gamma_j, 2-2\gamma_j, 2\gamma_j\} \cup \{0,1,2\}$. It is convenient to express log-likelihood and quasi-likelihood as follows. The likelihood satisfies

$$\ln \mathcal{L}_T(\theta;\sigma,\gamma) := \sum_{j=0}^{d} \frac{1}{\sigma_j^2}\left(a_j M_T^{j,-2\gamma_j} - b_j M_T^{j,1-2\gamma_j} - \frac{a_j^2}{2} Q_T^{j,-2\gamma_j} - a_j b_j Q_T^{j,1-2\gamma_j} - \frac{b_j^2}{2} Q_T^{j,2-2\gamma_j}\right), \tag{3.2}$$

and the quasi-likelihood $q\text{-}\mathcal{L}_T(\theta) := \ln \mathcal{L}_T(\theta;1,0)$, which rewrites as

$$q\text{-}\mathcal{L}_T(\theta) = \sum_{j=0}^{d} a_j M_T^{j,0} - b_j M_T^{j,1} - \frac{a_j^2}{2} Q_T^{j,0} - a_j b_j Q_T^{j,1} - \frac{b_j^2}{2} Q_T^{j,2}. \tag{3.3}$$

The following proposition provides explicit expression of MLE and QMLE, in terms of the quantities in (3.1).

Proposition 3.1. *Let $T \in (0,\infty)$, The maximum of the likelihood $\mathcal{L}_T(\theta;\sigma,\gamma)$ is achieved at $\theta_T^{(\mathcal{L})} := (a_T^{j,\gamma}, b_T^{j,\gamma})_{j=0}^d$ with*

$$(a_T^{j,\gamma}, b_T^{j,\gamma}) = \left(\frac{M_T^{j,-2\gamma_j} Q_T^{j,2-2\gamma_j} - Q_T^{j,1-2\gamma_j} M_T^{j,1-2\gamma_j}}{Q_T^{j,-2\gamma_j} Q_T^{j,2-2\gamma_j} - (Q_T^{j,1-2\gamma_j})^2}, \frac{M_T^{j,-2\gamma_j} Q_T^{j,1-2\gamma_j} - Q_T^{j,-2\gamma_j} M_T^{j,1-2\gamma_j}}{Q_T^{j,-2\gamma_j} Q_T^{j,2-2\gamma_j} - (Q_T^{j,1-2\gamma_j})^2}\right). \tag{3.4}$$

The maximum of the quasi-likelihood $q\text{-}\mathcal{L}_T(\theta)$ is achieved at $\theta_T^{(q\text{-}\mathcal{L})} := (a_T^{j,0}, b_T^{j,0})_{j=0}^d$, that is

$$(a_T^{j,0}, b_T^{j,0}) = \left(\frac{M_T^{j,0} Q_T^{j,2} - Q_T^{j,1} M_T^{j,1}}{Q_T^{j,0} Q_T^{j,2} - (Q_T^{j,1})^2}, \frac{M_T^{j,0} Q_T^{j,1} - Q_T^{j,0} M_T^{j,1}}{Q_T^{j,0} Q_T^{j,2} - (Q_T^{j,1})^2}\right). \tag{3.5}$$

Proof. We sketch the proof for MLE. The same works for QMLE. One shows that (3.4) is the unique singular point of the gradient (vector of the derivatives with respect to a_j and b_j for all $j \in \{0, \cdots, d\}$) of (3.2) and the Hessian is negative definite. □

Remark 3.2. *If $\gamma \equiv 0$, then the diffusion coefficient is piecewise constant, so T-CKLS is a threshold OU (T-OU) and QMLE and MLE coincide, as noticed in [24].*

Remark 3.3. *For every $j = 0, \ldots, d$, $(a_T^{j,\gamma}, b_T^{j,\gamma})$ only depend on the observations of the trajectory $t \mapsto X_t$ which belong to I_j.*

MLE and QMLE do not depend on $\sigma_\star$ explicitly, but only on the quantities in (3.1). The following result ensures that $\sigma_\star$ is a.s. equal to an estimator expressed in terms of $Q_T^{j,\cdot}, M_T^{\cdot,0}, M_T^{j,1}, X_T, X_0$.

Proposition 3.4. *Let $T \in (0,\infty)$ and $j \in \{0, \cdots, d\}$. Then*

$$\sigma_j = \sqrt{\frac{\mathbf{Q}_T^j}{Q_T^{j,2\gamma_j}}} \quad \text{a.s. on the event } \{Q_T^{j,0} > 0\} \tag{3.6}$$

where, for $j \in \{0, d\}$,

$$\mathbf{Q}_T^j := (f_j(X_T))^2 - (f_j(X_0))^2 + 2\left(r_{(j+1)\wedge d} M_T^{j,0} - M_T^{j,1}\right) \tag{3.7}$$

with $f_j(x) = \mathbb{1}_{I_j}(x)|r_j - x|$, *otherwise when* $j \in \{1, \dots, d-1\}$*:*

$$\mathbf{Q}_T^j := (f_j(X_T))^2 - (f_j(X_0))^2 - 2M_T^{j,1} + 2\left(r_j \sum_{i=j}^{d} M_T^{i,0} - r_{j+1} \sum_{i=j+1}^{d} M_T^{i,0}\right) \tag{3.8}$$

with $f_j(x) = \mathbb{1}_{I_j}(x)(x - r_j) + r_{j+1}\mathbb{1}_{[r_{j+1}, +\infty)}(x)$.

Proof. We only consider the case $j \in \{1, \dots, d-1\}$. When $j \in \{0, d\}$ the proof works analogously. Considering the event $\{Q_T^{j,0} > 0\}$ corresponds to take trajectories which spend some time in I_j and so $Q_T^{j,2\gamma_j}$ does not vanish. Applying Itô-Tanaka formula (see [29, Chapter VI, exercice 1.25]) ensures that

$$\mathrm{d}f_j(X_s) = (a_j - b_j X_s)\mathbb{1}_{I_j}(X_s)\,\mathrm{d}s + \sigma_j(X_s)^{\gamma_j}\mathbb{1}_{I_j}(X_s)\,\mathrm{d}B_s + \frac{1}{2}\,\mathrm{d}\left(L_{r_j}^s(X) - L_{r_{j+1}}^s(X)\right).$$

The quadratic variation of $f_{j,1}(X)$ satisfies a.s. the equality:

$$\langle f_j(X)\rangle_T = \sigma_j^2 \int_0^T (X_s)^{2\gamma_j}\mathbb{1}_{I_j}(X_s)\,\mathrm{d}s = \sigma_j^2 Q_T^{j,2\gamma_j}.$$

Itô formula applied to $(f_j(X_T))^2$, ensures that $\langle f_j(X)\rangle_T$ is a.s. equal to Q_T^j, indeed for every semi-martingale Y it holds a.s. that $\mathrm{d}Y^2 = 2Y\,\mathrm{d}Y + \mathrm{d}\langle Y\rangle$. □

3.2 Asymptotic properties: long time

In this section, we explore the statistical properties as $T \to \infty$ of the MLE and QMLE from continuous time observations of a trajectory of the T-CKLS process.
We assume that the process is ergodic, μ is the stationary distribution given in Section 7.1, and we introduce the following hypotheses:

- $\mathbf{H}_{\mathcal{L}}$: μ admits finite $(-2\max_{j=0,\cdots,d}\gamma_j)$-th and $(2 - 2\max_{j=0,\cdots,d}\gamma_j)$-th moment,
- $\mathbf{H}_{q\text{-}\mathcal{L}}$: μ admits finite $(2 + 2\max_{j=0,\cdots,d}\gamma_j)$-th moment.

See Section 5.3 for comments on these assumptions.
The asymptotic behavior of the MLE and QMLE are provided in the following theorem which states that MLE and QMLE are strongly consistent and asymptotically normal estimators of the drift parameter $\theta_\star$.

Theorem 3.5. *For* $\ell \in \{\mathcal{L}, q\text{-}\mathcal{L}\}$, *under Hypothesis* $\mathbf{H}_\ell$, *the MLE and QMLE are strongly consistent estimators of* $\theta_\star$ *i.e.*

$$\theta_T^{(\ell)} \xrightarrow[T\to+\infty]{a.s.} \theta_\star. \tag{3.9}$$

Furthermore, the following convergence is satisfied:

$$\sqrt{T}\left(\theta_T^{(\ell)} - \theta_\star\right) \xrightarrow[T\to+\infty]{stably} \mathbf{N}^{(\ell)}, \tag{3.10}$$

where $\mathbf{N}^{(\ell)} = (N^{(\ell)}_{j,a}, N^{(\ell)}_{j,b})_{j=0}^{d}$ *are* $d+1$ *independent, independent of* X*, two-dimensional centered Gaussian random variables with covariance matrices respectively given by* $\sigma_j^2\Gamma_j^{(\ell)}$ *such that*

$$\Gamma_j^{(\mathcal{L})} := \Gamma_j^{(\mathcal{L},\gamma_j)} := \begin{pmatrix} Q_\infty^{j,-2\gamma_j} & -Q_\infty^{j,1-2\gamma_j} \\ -Q_\infty^{j,1-2\gamma_j} & Q_\infty^{j,2-2\gamma_j} \end{pmatrix}^{-1} \quad \textit{and} \quad \Gamma_j^{(q\text{-}\mathcal{L})} = \Gamma_j^{(\mathcal{L},0)}\Gamma_j^{(\mathcal{L},-\gamma_j)}\Gamma_j^{(\mathcal{L},0)}$$

where $Q_\infty^{j,\cdot}$ *are real constant defined in Lemma 3.6.*

Proof. The proof is similar to the one in [22, 24], nevertheless we summarize the steps and stress the specificity of the case we consider in this article. Note that MLE and QMLE rewrite as follows:

$$\theta_j^{(\mathcal{L})} \overset{a.s.}{=} \theta_j + \sigma_j \left(\mathfrak{M}_T^{j,-\gamma_j}, -\mathfrak{M}_T^{j,1-\gamma_j}\right)\Gamma^{(\mathcal{L})} \quad \text{and} \quad \theta_j^{(q\text{-}\mathcal{L})} \overset{a.s.}{=} \theta_j + \sigma_j \left(\mathfrak{M}_T^{j,\gamma_j}, -\mathfrak{M}_T^{j,1+\gamma_j}\right)\Gamma^{(\mathcal{L},0)}$$

where $\mathfrak{M}_T^{j,k} = \int_0^T (X_s)^k \mathbb{1}_{I_j}(X_s)\,\mathrm{d}B_s$ for $k \in \{-\gamma_j, 1-\gamma_j, \gamma_j, 1+\gamma_j\}$ are martingales. Indeed, note that $Q_T^{j,2k+2\gamma_j}$ is the quadratic variation of $\mathfrak{M}_T^{j,k}$ which is the one of $M_T^{j,k}$ up to a multiplicative factor. We can now exploit martingale theorems. The consistency of the estimator (MLE and QMLE) follows directly from [23, Theorem 1] and the ergodicity of the process which implies, for instance, Lemma 3.6. The asymptotic normality property follows from [9, Theorem 2.2]. Hypotheses $\mathbf{H}_{\mathcal{L}}$ and $\mathbf{H}_{q\text{-}\mathcal{L}}$ are necessary for the application of [23, Theorem 1]. □

Lemma 3.6 (Ergodic properties)**.** *For* $j \in \{0, \cdots, d\}$ *and* $m \in \mathbb{R}$*, if the* m*-th moment of* μ *is finite on the set* I_j*, then*

$$Q_\infty^{j,m} \overset{a.s.}{:=} \lim_{T\to\infty} \frac{Q_T^{j,m}}{T} = \int_{I_j} x^m \mu(\,\mathrm{d}x),$$

are non-vanishing constants.

Remark 3.7. *The above asymptotic normality of the estimators implies the local asymptotic normality (LAN) property (see [18]). The LAN property, is a fundamental concept in the asymptotic theory of statistics. For instance, when it is satisfied, it can be combined with the Minimax theorem to establish a lower bound for the asymptotic variance of estimators.*
Since statement and proof are analogous to the one of the T-OU process (corresponding to $\gamma = 0$*) given in [24, Theorem 1.(iv)], we just provide a short statement.*
Let $\ell \in \{\mathcal{L}, q\text{-}\mathcal{L}\}$ *and assume that* $\mathbf{H}_\ell$ *holds. The LAN property holds for the* ℓ*-function with rate of convergence* $1/\sqrt{T}$*. Furthermore, the asymptotic Fisher information is given by*

$$\Gamma^{(\ell)} = \begin{bmatrix} \frac{1}{\sigma_0^2}\left(\Gamma_0^{(\ell)}\right)^{-1} & 0_{\mathbb{R}^{2\times 2}} & \dots & 0_{\mathbb{R}^{2\times 2}} \\ 0_{\mathbb{R}^{2\times 2}} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \frac{1}{\sigma_{d-1}^2}\left(\Gamma_{d-1}^{(\ell)}\right)^{-1} & 0_{\mathbb{R}^{2\times 2}} \\ 0_{\mathbb{R}^{2\times 2}} & \dots & 0_{\mathbb{R}^{2\times 2}} & \frac{1}{\sigma_d^2}\left(\Gamma_d^{(l)}\right)^{-1} \end{bmatrix}.$$

4 Estimation from discrete observations

In this section, we assume to observe the process on a discrete time grid $0 = t_0 < t_1 \cdots < t_{N-1} < t_N = T_N < \infty$, for $N \in \mathbb{N}$.
First, we provide estimators which maximize a discretized versions of likelihood $\mathcal{L}$ and quasi-likelihood q-$\mathcal{L}$. Next, we study the asymptotic behavior of the estimators in high frequency and long time under the assumption that the process is stationary.

4.1 Estimators expressions

There is no exploitable explicit expression for the transition densities of the T-CKLS process, nor for the finite dimensional distributions. This is true also in the well known special case of threshold Brownian motion with piecewise constant drift. Hence, instead of considering a likelihood or quasi-likelihood function, we considered a discretization of the likelihood $\mathcal{L}_T$ (2.4) and quasi-likelihood q-$\mathcal{L}_T$ (2.5). Further comments on this choice are given in Section 5. Once these discretizations introduced, we denote them respectively by $\mathcal{L}_{T_N,N}$ and q-$\mathcal{L}_{T_N,N}$ and we compute the estimators

$$\theta^{(\mathcal{L})}_{T_N,N} = \underset{\theta\in\Theta^{(\mathcal{L})}}{\operatorname{argmax}} \mathcal{L}_{T_N,N}(\theta;\sigma,\gamma) \quad \text{and} \quad \theta^{(q\text{-}\mathcal{L})}_{T_N,N} = \underset{\theta\in\Theta^{(q\text{-}\mathcal{L})}}{\operatorname{argmax}}\ q\text{-}\mathcal{L}_{T_N,N}(\theta). \tag{4.1}$$

Let us denote by $Q^{j,m}_{T_N,N}$ the discrete versions of $Q^{j,m}_{T_N,N}$ and $M^{j,m}_{T_N,N}$ in (3.1):

$$Q^{j,m}_{T_N,N} := \sum_{i=0}^{N-1} X^m_{t_i} \mathbb{1}_{I_j}(X_{t_i})(t_{i+1} - t_i) \quad \text{and} \quad M^{j,m}_{T_N,N} := \sum_{i=0}^{N-1} X^m_{t_i} \mathbb{1}_{I_j}(X_{t_i})(X_{t_{i+1}} - X_{t_i}) \tag{4.2}$$

for $j \in \{0, \cdots, d\}$ and $m \in \{-2\gamma_j, 1-2\gamma_j, 2-2\gamma_j, 2\gamma_j\} \cup \{-1, 0, 1, 2\}$. To obtain the discretized (quasi-)likelihood, it would be natural to replace the above quantities in the continuous-time observations (quasi-)likelihood function given in (3.2)-(3.3). The discretized quasi-likelihood is then

$$q\text{-}\mathcal{L}_{T_N,N}(\theta) := \sum_{j=0}^{d} a_j M^{j,0}_{T_N,N} - b_j M^{j,1}_{T_N,N} - \frac{a_j^2}{2} Q^{j,0}_{T_N,N} - a_j b_j Q^{j,1}_{T_N,N} - \frac{b_j^2}{2} Q^{j,2}_{T_N,N}, \tag{4.3}$$

and we *could* do similarly for the likelihood. We would get

$$\ln \mathcal{L}_{T_N,N}(\theta;\sigma,\gamma) = \sum_{j=0}^{d} \frac{1}{\sigma_j^2} \left(a_j M^{j,-2\gamma_j}_{T_N,N} - b_j M^{j,1-2\gamma_j}_{T_N,N} - \frac{a_j^2}{2} Q^{j,-2\gamma_j}_{T_N,N} - a_j b_j Q^{j,1-2\gamma_j}_{T_N,N} - \frac{b_j^2}{2} Q^{j,2-2\gamma_j}_{T_N,N} \right). \tag{4.4}$$

Actually, we do not choose the latter quantity as discretized-likelihood. Instead, we consider a different discretization of $M^{j,m}$ based on an alternative expression which depends on $M^{j,0}$ and $Q^{j,0}$ and $Q^{j,-1}$ (see Lemma 5.1 in Section 5, where we also explain this choice for the discretization). For every $j \in \{0, \ldots, d\}$ we define $\mathcal{M}^{j,0}_{T_N,N} := M^{j,0}_{T_N,N}$ and for $m \in \{-2\gamma_j, 1-2\gamma_j\} \setminus \{0\}$,

$$\begin{aligned} \mathcal{M}^{j,m}_{T_N,N} := {} & f_{j,m+1}(X_{T_N}) - f_{j,m+1}(X_0) - \frac{m}{2}\sigma_j^2 Q^{m+2\gamma_j-1}_{T_N,N} \\ & + \sum_{k=j}^{j+1} (-1)^{1-k} r_k^m \sum_{i=k}^{d} \left(M^{i,0}_{T_N,N} + \sum_{l=k}^{d} f_{l,1}(X_0) - f_{l,1}(X_{T_N}) \right), \end{aligned} \tag{4.5}$$

if $j \in \{1, \dots, d-1\}$ and if $j \in \{0, d\}$:

$$\begin{aligned}\mathcal{M}_T^{j,m} :=& f_{j,m+1}(X_{T_N}) - f_{j,m+1}(X_0) - \frac{m}{2}\sigma_j^2 Q_{T_N,N}^{m+2\gamma_j-1} \\ &+ (r_{(j+1)\wedge d})^m \left[M_{T_N,N}^{j,0} + (f_{j,1}(X_0) - f_{j,1}(X_{T_N}))\right], \end{aligned} \tag{4.6}$$

where $f_{0,m}(x) = \max\{0, \int_x^{r_1} y^{m-1}dy\}$, $f_{d,m}(x) = \max\{0, \int_{r_d}^x y^{m-1}dy\}$ and for $j \in \{1, \cdots, d-1\}$, $f_{j,m}(x) = \max\{0, \int_{r_j}^x y^{m-1}dy\} - \max\{0, \int_{r_{j+1}}^x y^{m-1}dy\}$. Finally, we consider the following discretized log-likelihood:

$$\ln \mathcal{L}_{T_N,N}(\theta; \sigma, \gamma) = \sum_{j=0}^{d} \frac{1}{\sigma_j^2}\left(a_j \mathcal{M}_{T_N,N}^{j,-2\gamma_j} - b_j \mathcal{M}_{T_N,N}^{j,1-2\gamma_j} - \frac{a_j^2}{2} Q_{T_N,N}^{j,-2\gamma_j} - a_j b_j Q_{T_N,N}^{j,1-2\gamma_j} - \frac{b_j^2}{2} Q_{T_N,N}^{j,2-2\gamma_j}\right). \tag{4.7}$$

The following proposition establishes an explicit expressions of the discretized MLE and discretized QMLE.

Proposition 4.1. *Let $(T_N)_{N\in\mathbb{N}}$ be a sequence in $(0,\infty)$ and let $N \in \mathbb{N}$. The maximum of the discretized likelihood is achieved at $\theta_{T_N,N}^{(\mathcal{L})} = (a_{T_N,N}^{j,\gamma_j}, b_{T_N,N}^{j,\gamma_j})_{j=0}^d$ with*

$$(a_{T_N,N}^{j,\gamma_j}, b_{T_N,N}^{j,\gamma_j}) := \left(\frac{\mathcal{M}_{T_N,N}^{j,-2\gamma_j} Q_{T_N,N}^{j,2-2\gamma_j} - Q_{T_N,N}^{j,1-2\gamma_j} \mathcal{M}_{T_N,N}^{j,1-2\gamma_j}}{Q_{T_N,N}^{j,-2\gamma_j} Q_{T_N,N}^{j,2-2\gamma_j} - (Q_{T_N,N}^{j,1-2\gamma_j})^2}, \frac{\mathcal{M}_{T_N,N}^{j,-2\gamma_j} Q_{T_N,N}^{j,1-2\gamma_j} - Q_{T_N,N}^{j,-2\gamma_j} \mathcal{M}_{T_N,N}^{j,1-2\gamma_j}}{Q_{T_N,N}^{j,-2\gamma_j} Q_{T_N,N}^{j,2-2\gamma_j} - (Q_{T_N,N}^{j,1-2\gamma_j})^2}\right). \tag{4.8}$$

The maximum of discretized quasi-likelihood is achieved at $\theta_{T_N,N}^{(q\text{-}\mathcal{L})} = (a_{T_N,N}^{j,0}, b_{T_N,N}^{j,0})_{j=0}^d$ with

$$(a_{T_N,N}^{j,0}, b_{T_N,N}^{j,0}) = \left(\frac{M_{T_N,N}^{j,0} Q_{T_N,N}^{j,2} - Q_{T_N,N}^{j,1} M_{T_N,N}^{j,1}}{Q_{T_N,N}^{j,0} Q_{T_N,N}^{j,2} - (Q_{T_N,N}^{j,1})^2}, \frac{M_{T_N,N}^{j,0} Q_{T_N,N}^{j,1} - Q_{T_N,N}^{j,0} M_{T_N,N}^{j,1}}{Q_{T_N,N}^{j,0} Q_{T_N,N}^{j,2} - (Q_{T_N,N}^{j,1})^2}\right). \tag{4.9}$$

The proof is omitted because it is analogous to the one of Proposition 3.1.
Note that the QMLE does not depend on the parameter vector $\sigma_\star = (\sigma_j)_{j=0}^d$, instead the MLE does because the expressions $\mathcal{M}_{T_N,N}^{j,m}$ involve the parameter $\sigma_\star$. Since we assume σ is not known, we replace it by an estimator.
For $j \in \{0, \cdots, d\}$,

$$\sigma_{T_N,N}^j = \sqrt{\frac{\mathbf{Q}_{T_N,N}^j}{Q_{T_N,N}^{j,2\gamma_j}}}, \tag{4.10}$$

where $\mathbf{Q}_{T_N,N}^j$ is obtained by discretizing the right hand side of formula (3.7) and (3.8) in Proposition 3.4. Note that $\mathbf{Q}_{T_N,N}^j$ depends on $M_{T_N,N}^{j,1}$ and $M_{T_N,N}^{j,0}$ defined in (4.2). For instance,

$$\mathbf{Q}_{T_N,N}^d := (f_d(X_{T_N}))^2 - (f_d(X_0))^2 + 2\left(r_d M_{T_N,N}^{d,0} - M_{T_N,N}^{d,1}\right).$$

4.2 Asymptotic properties: high frequency - long time

In this section, we state the statistical properties of the discretized MLE and QMLE.
Let $\Delta_N := \max_{k=1,\cdots,N}\{t_k - t_{k-1}\}$ denote the maximal lag between two consecutive observations. We assume that the observation time window goes to infinity (long time) and the maximal lag between consecutive observations vanishes (high frequency):

$$\lim_{N\to+\infty} T_N = +\infty \quad \text{and} \quad \lim_{N\to+\infty} \Delta_N = 0. \tag{4.11}$$

Moreover we assume that the process is stationary.
The results of this section require additional assumptions on the moments of the stationary distribution:

- $\mathcal{H}_{\mathcal{L}}$: μ admits finite $(-p)$-th and 4-th moments with $p > 4 \max_{j=0,\cdots,d} \gamma_j$.
- $\mathcal{H}_{q\text{-}\mathcal{L}}$: μ admits finite 4-th moment.

These assumptions depend on the fact that $\sigma_\star$ is supposed to be unknown and on the discretization chosen for the likelihood and quasi-likelihood as well. See Section 5.3 for more details.
Without loss of generality we assume $\Delta_N \in (0,1)$ for all N. Then, we also introduce the following quantities $g_N^{(q\text{-}\mathcal{L})} = \Delta_N$,

$$g_N^{(\mathcal{L})} = \max_{j=0,\cdots,d} \begin{cases} \Delta_N^{2-2\gamma_j} & \text{if } \gamma_j \in (\frac{3}{4}, 1), \\ \Delta_N^{2\gamma_j - 1} & \text{if } \gamma_j \in (\frac{1}{2}, \frac{3}{4}], \\ \Delta_N^{1-2\gamma_j} & \text{if } \gamma_j \in (\frac{1}{4}, \frac{1}{2}), \\ \Delta_N^{2\gamma_j} & \text{if } \gamma_j \in (0, \frac{1}{4}], \\ \Delta_N & \text{if } \gamma_j \in \{0, \frac{1}{2}, 1\}, \end{cases} \qquad g_N^{(\sigma)} := \max_{j=0,\cdots,d} \begin{cases} \Delta_N^{\gamma_j} & \text{if } \gamma_j \in (\frac{1}{2}, 1), \\ \Delta_N^{2\gamma_j} & \text{if } \gamma_j \in (0, \frac{1}{2}), \\ \Delta_N & \text{if } \gamma_j \in \{0, \frac{1}{2}, 1\}. \end{cases} \tag{4.12}$$

Remark 4.2. *If* $\max\{\gamma_j \colon j = 0, \ldots, d\} \in \{0, 1/2, 1\}$, *then* $\Delta_N = g_N^{(q\text{-}\mathcal{L})} = g_N^{(\mathcal{L})} = g_N^{(\sigma)}$.

We are now ready to provide our first convergence result in the discrete setting. We consider the volatility estimator in (4.10) and we prove consistency, and show that the speed of convergence is higher than $\sqrt{T_N}$.

Theorem 4.3. *Assume that* (4.11) *holds, that the T-CKLS X is stationary and that Hypothesis* $\mathcal{H}_{q\text{-}\mathcal{L}}$ *holds. Then, the estimator* $\sigma^2_{T_N,N} = ((\sigma^j_{T_N,N})^2)_{j=0}^d$ *in* (4.10) *is a consistent estimators of the diffusion coefficient vector* $\sigma^2_\star = ((\sigma_j)^2)_{j=0}^d$, *i.e.*

$$\sigma^2_{T_N,N} \xrightarrow[N\to+\infty]{\mathbb{P}} \sigma^2_\star. \tag{4.13}$$

Under the additional assumption that $\lim_{N\to+\infty} T_N g_N^{(\sigma)} = 0$, *it holds that*

$$\sqrt{T_N}\left(\sigma^2_{T_N,N} - \sigma^2_\star\right) \xrightarrow[N\to+\infty]{\mathbb{P}} 0_{\mathbb{R}^{d+1}}. \tag{4.14}$$

Since we assume that the diffusion coefficient σ vector is unknown, we replace it by estimator (4.10) in the expression of $\mathcal{M}^{j,\cdot}_{T_N,N}$.

Remark 4.4. *The results of the section about MLE and QMLE estimation hold with less restrictive assumptions if* $\sigma_\star$ *is known, see Section 5.3.*

The following theorems state the asymptotic properties in high frequency and long time observations of the discretized MLE and QMLE of $\theta_\star$.

Theorem 4.5. *Assume that* (4.11) *holds and that the T-CKLS X is stationary. For* $\ell \in \{\mathcal{L}, q\text{-}\mathcal{L}\}$, *under Hypothesis* $\mathcal{H}_\ell$, *the MLE and QMLE are weak consistent estimators of* $\theta_\star$ *i.e.*

$$\theta^{(\ell)}_{T_N,N} \xrightarrow[N\to+\infty]{\mathbb{P}} \theta_\star. \tag{4.15}$$

Furthermore, if $\lim_{N\to+\infty} T_N g_N^{(\ell)} = 0$*, then under the same hypothesis, we have*

$$\sqrt{T_N}\left(\theta^{(\ell)}_{T_N,N} - \theta_\star\right) \xrightarrow[N\to+\infty]{stably} \mathbf{N}^{(\ell)}, \tag{4.16}$$

where $\mathbf{N}^{(.)}$ *is defined in Theorem 3.5.*

The proof of Theorem 4.3 and Theorem 4.5 are postponed to Section 7. Nevertheless we give the main ideas and tools in Section 5.

Remark 4.6. *For* $\ell \in \{\mathcal{L}, q\text{-}\mathcal{L}\}$*, it follows from the previous results that the estimator* $(\theta^{(\ell)}_{T_N,N}, \sigma_{T_N,N})$ *is a consistent estimator of* $(\theta_\star, \sigma_\star)$*. By the stable convergence properties,* $\sqrt{T_N}(\theta^{(\ell)}_{T_N,N}, \sigma_{T_N,N})$ *converges stably to the vector* $(\mathbf{N}^{(\ell)}, 0_{\mathbb{R}^{d+1}})$ *If* σ *is known, the analogous to the LAN property in Remark 3.7 holds for the discretized MLE and QMLE.*

5 Comments on the results and their proofs

In this section, we comment the results of the previous section and we summarize the key elements of the proofs of consistency and asymptotic normality of the continuous time and discretized MLE and QMLE. Moreover, we discuss assumptions and extensions, and compare with related literature.
The main results of this article are the asymptotic results in the context of a T-CKLS process observed in high frequency and long time: Theorems 4.3 and 4.5. As we already mentioned, these results are new also in the context of other threshold diffusions such as T-CIR, and for mixed dynamics: CIR on a space interval, OU on another, CKLS on another one. We exploit this feature in Section 6. In Section 5.1 we comment on the estimators. In particular, we discuss the discretization choice $\mathcal{M}$ and the novelty on estimation of $\sigma_\star$.
Since the process is ergodic, the proofs of the asymptotic results in the context of continuous time observations follow from standard martingale central limit theorems. This is the case in other results for TDs, e.g. [30, 22, 24]. Therefore, in this section, we only comment on the proofs in the case of discrete observations. This is done in Section 5.2. Nevertheless, in Section 5.3, we provide some comments on the assumptions of the results both in the context of discrete and continuous time observations and we compare our assumptions with those considered in the literature.

5.1 On the estimators

On the volatility estimator. Estimator (4.10) is inspired by the results in [20]. The latter reference studies some estimators for the volatility of oscillating Brownian motion (null drift, $\gamma = 0$, one threshold) from high frequency discrete observations over a finite time horizon. The estimators are based on quadratic variation, but on two separate intervals over which the volatility is constant. Our estimator exploits a different discretization choice for the quadratic variation over the intervals I_j, which allows to obtain information about its behavior in high frequency and long time. Up to our knowledge, this is the first time an estimator of the volatility for TDs (even with $\gamma \equiv 0$) is analyzed in high frequency and long time.

Likelihood discretization: $M^{\cdot,m}_{T_N,N}$ versus $\mathcal{M}^{\cdot,m}_{T_N,N}$. Let us first introduce an expression $\mathbb{P}$-a.s. equal to $M^{j,m}_T$ which would not involve any term $M^{j,k}$ except if $k=0$.

Lemma 5.1. *Let $T\in(0,\infty)$, $j\in\{1\cdots,d\}$ and $m\in\{-2\gamma_j,1-2\gamma_j\}$. It holds $\mathbb{P}$-a.s. that*

$$M^{j,m}_T = f_{j,m+1}(X_T) - f_{j,m+1}(X_0) - \frac{m}{2}\sigma_j^2 Q_T^{m+2\gamma_j-1} + \sum_{k=j}^{j+1}(-1)^{1-k} r_k^m \sum_{i=k}^{d}\left(M^{i,0}_T + \sum_{l=k}^{d} f_{l,1}(X_0) - f_{l,1}(X_T)\right), \tag{5.1}$$

for $j\in\{1,\dots,d-1\}$ and for $j\in\{0,d\}$:

$$M^{j,m}_T = f_{j,m+1}(X_T) - f_{j,m+1}(X_0) - \frac{m}{2}\sigma_j^2 Q_T^{m+2\gamma_j-1} + (r_{(j+1)\wedge d})^m\left[M^{j,0}_T + (f_{j,1}(X_0) - f_{j,1}(X_T))\right]. \tag{5.2}$$

Proof. We prove only the case $j\in\{1,\dots,d-1\}$, and the case $j\in\{0,d\}$ works similarly. Applying Itô-Tanaka formula (see e.g. [29, Chapter VI, exercice 1.25]), yields the a.s. equality

$$f_{j,m+1}(X_T) = f_{j,m+1}(X_0) + \frac{m}{2}\sigma_j^2 Q^{m+2\gamma_j-1} + M^{j,m}_T + \frac{r_j^m}{2}\left(L_T^{r_j}(X) - L_T^{r_{j+1}}(X)\right). \tag{5.3}$$

We now replace the dependence on the local time on a dependence on $M^{j,0}, X_T, X_0$. Applying a second time Itô-Tanaka formula to $\sum_{i=k}^d f_{i,0}(X_T)$ with $k\in\{1,\dots,d-1\}$, shows that a.s.

$$\frac{1}{2}L_T^{r_k}(X) = \sum_{i=k}^{d} f_{i,0}(X_T) - f_{i,0}(X_0) - M^{i,0}_T. \tag{5.4}$$

Combining equation (5.4) for $k=j$ and $k=j+1$ with (5.3) completes the proof. □

We used the latter result to obtain $\mathcal{M}^{j,m}_{T_N,N}$: we just considered the discretized versions of the right hand side of (5.1) and (5.2) in Lemma 5.1 by replacing the quantities $M^{j,0}$ and $Q^{j,m+2\gamma_j-1}$ by their discretized versions (4.2).
This choice impacts hypothesis $\mathcal{H}_{\mathcal{L}}$. See Section 5.3 for more details.

5.2 Key elements of proofs, comments and extensions.

We now comment on the proofs of consistency and asymptotic normality for drift and diffusion coefficient, namely Theorems 4.3 and 4.5. They rely on the next Lemma 5.2 and on the continuous time results Proposition 3.4 for the drift in Theorem 3.5. More precisely, one shows that the rescaled difference between the discrete-time and continuous-time estimators (e.g. $\sqrt{T_N}\left(\theta^{(\mathcal{L})}_{T_N,N} - \theta^{(\mathcal{L})}_{T_N}\right)$) vanishes as $N\to\infty$. Since all estimators depend on $M^{j,m}_{T_N,N}$ and $Q^{j,m}_{T_N,N}$, the proof is based on the following result.

Lemma 5.2. *Assume that (4.11) holds and that the T-CKLS X is stationary. Let $\lambda\in\{1,2\}$, $j\in\{0,\cdots,d\}$, and $m\in\{0,1\}$. Then*

$$\lim_{N\to\infty} T_N^{-1/\lambda}\mathbb{E}\left[|Q^{j,k}_{T_N} - Q^{j,k}_{T_N,N}|\right] = 0 \quad \textit{and} \quad \lim_{N\to\infty} T_N^{-1/\lambda}\mathbb{E}\left[|M^{j,m}_{T_N} - M^{j,m}_{T_N,N}|\right] = 0 \tag{5.5}$$

in each one of the following cases

1. *for* $k \in \{2\gamma_j\}$,
   under the assumptions: $\lim_{N\to+\infty} T_N^{1-1/\lambda} g_N^{(\sigma)} = 0$ *and hypothesis* $\mathcal{H}_{q\text{-}\mathcal{L}}$ *holds.*

2. *for* $k \in \{-2\gamma_j, 1-2\gamma_j, 2-2\gamma_j, 2\gamma_j\} \cup \{-1, 0, 1\}$,
   under the assumptions: $\lim_{N\to+\infty} T_N^{1-1/\lambda} g_N^{(\mathcal{L})} = 0$ *and hypothesis* $\mathcal{H}_{\mathcal{L}}$ *holds,*

3. *for* $k \in \{0, 1, 2\}$,
   under the assumptions: $\lim_{N\to+\infty} T_N^{1-1/\lambda} g_N^{(q\text{-}\mathcal{L})} = 0$ *and hypothesis* $\mathcal{H}_{q\text{-}\mathcal{L}}$ *holds.*

The proof of Lemma 5.2 is provided in Section 7. It is quite technical so we give the main ideas and tools here.
The proof of Lemma 5.2 relies on two auxiliary results for T-CKLS processes: Propositions 5.3-5.4 whose proofs are postponed to Section 7.5 and Section 7.6 respectively. Proposition 5.3 is a property very commonly exploited in statistics for diffusion processes.

Proposition 5.3. *Assume that the T-CKLS X is stationary. Let $j \in \{0, \cdots, d\}$ and $m \geq 1$. Assume that the m-th moment of μ is finite. Then there exists a constant $C \in (0, \infty)$ such that for all $0 \leq s < t$ it holds $\mathbb{E}\left[|X_t - X_s|^m\right] \leq C(t-s)^{m/2}$.*

Proposition 5.4, instead, is the remedy to the lack of knowledge of the finite dimensional distributions of TDs. Indeed, the quantities $M^{j,\cdot}, Q^{j,\cdot}$ consider only observations taking values on I_j together with their following observation. Hence, in the proof of Lemma 5.2, one needs to bound the probability that the process crossed a threshold between two consecutive observations.

Proposition 5.4. *Assume that the T-CKLS X is stationary. Let $j \in \{0, \cdots, d\}$ and $m \in \mathbb{R}$, if the m-th moment of μ is finite on the set I_j. Then there exists a constant $C \in (0, \infty)$ such that for all $0 \leq s < t$, we have:*

$$\mathbb{E}\left[|X_s|^m \mathbb{P}_{X_s}\left(\tau_{I_j}^{\xi^j} < t - s\right) \mathbb{1}_{I_j}(X_s)\right] \leq C(t-s)^{1/2}, \tag{5.6}$$

where ξ^j is a CKLS process with parameters $(a_j, b_j, \sigma_j, \gamma_j)$ starting at X_s and driven by a Brownian motion independent of $\mathcal{F}_s$, and $\tau_{I_j}^{\xi^j}$ is the first hitting time of the interval I_j for the process ξ^j.

Differences with respect to related literature Let us first consider standard diffusions related to T-CKLS, for instance CIR process. The proofs in [4, 5] rely on the knowledge of the law of the process, and some other quantities such as the integral $\int_0^t (X_s)^{-1} \mathrm{d}s$. We do not have access to the law of T-CKLS, nor even of T-Brownian motion. Under the ergodicity assumption, in the context of discrete observations our Theorem 4.5, recovers and improves the existing result for CIR.
The proof strategy that we just illustrated above, is analogous to the one exploited in [24] for drift estimation of T-OU. Nevertheless, we believe that the way we deal with controlling the probability of crossing a threshold between two consecutive observations (proof of Lemma 5.4) is the key to extend the results of this paper to more general TDs. Hence, in our opinion, Proposition 5.4 is one of the most important results of this document.

5.3 Comments on assumptions

Assumptions in the continuous setting: $\mathbf{H}_{q\text{-}\mathcal{L}}$ and $\mathbf{H}_{\mathcal{L}}$. In the case of T-OU process (solution to 2.1 with $\gamma \equiv 0$), $\mathbf{H}_{q\text{-}\mathcal{L}}$ and $\mathbf{H}_{\mathcal{L}}$ rewrite as follows: μ admits finite second moment. When T-OU is ergodic (see Table 4), μ admits moments of all order (see Proposition A.3), so $\mathbf{H}_{q\text{-}\mathcal{L}}$ and $\mathbf{H}_{\mathcal{L}}$ hold. Indeed, in [24], no additional assumptions have been introduced.
For CIR process (solution to the SDE (1.1), with $\gamma = 1/2$). The asymptotic behavior of the MLE of an ergodic standard CIR process with $a, b > 0$ is studied in [5] under the additional condition $a \geq \sigma^2/2$. This condition ensures that μ admits finite (-1)-moment (see Proposition A.3) and so $\mathbf{H}_{\mathcal{L}}$ is satisfied.
In a similar way, in [26], the author study the asymptotic behavior of the MLE of a CKLS process, solution to the SDE (1.1), in the ergodic case with $a, b > 0$ and $\gamma \in (1/2, 1)$. Under these hypotheses on the parameters, the stationary distribution of the CKLS process admits moments of all order (see [26, Proposition 2.1] or next Proposition A.3). Therefore, $\mathbf{H}_{q\text{-}\mathcal{L}}$ and $\mathbf{H}_{\mathcal{L}}$ are satisfied.
We improve the conditions in [30], where consistency and asymptotic normality of the QMLE has been proven. More precisely, in [30], the process is supposed to be stationary and geometrically ergodic and they require finiteness of the fourth-moment. We only assume ergodicity, and, for the QMLE, assumption $H_{q\text{-}\mathcal{L}}$ is less restrictive than the existence of the fourth-order moment.

Assumptions in the discrete setting: the discretization choice. Let us note that in the likelihood function (4.4) there could be a term such as $M_T^{0,-1}$ (take $\gamma_0 = 1/2$). If one takes $M_{T,N}^{0,-1}$ for the discretization of the likelihood function, then one should prove the convergence in Lemma 5.2 for it. This requires stronger hypothesis than $\mathcal{H}_{\mathcal{L}}$, which instead is sufficient considering the discretization $\mathcal{M}_{T_N,N}^{0,-1}$ because, in this case, one only needs to prove Lemma 5.2 for the quantities $Q_{T_N,N}^{0,-1}$ and $M_{T_N,N}^{0,0}$.
We chose to illustrate the effect by taking $\gamma_0 = 1/2$, because the idea of considering a different expression is inspired by [5] which deals with non-threshold CIR process.

Assumptions in the discrete setting: asymptotic normality. In case $\gamma \equiv 0$ (T-OU) or $\gamma \equiv \frac{1}{2}$ (T-CIR), the hypotheses $\lim_{N\to\infty} T_N \Delta_N = 0$ has been considered for the asymptotic normality of estimators for CIR and T-OU process in [5, 24]. It has recently been considered also in [25] for T-drifted Brownian motion. The latter case is not covered by our article because we assume that $a > 0$. Our conditions for the convergence speed, $\lim_{N\to\infty} T_N g_N^{(\mathcal{L})} = 0$ and $\lim_{N\to\infty} T_N g_N^{(\sigma)} = 0$ degrade when the diffusion coefficient is non-linear in some interval, *i.e.* when $\max\{\gamma_j : j = 0, 1, \ldots, d\}$ is not equal to $0, 1/2$, or 1. Indeed, $\Delta_N = g_N^{(q\text{-}\mathcal{L})} \leq g_N^{(\mathcal{L})} \in \{\Delta_N\} \cup [\sqrt{\Delta}_N, 1)$ and $\Delta_N = g_N^{(q\text{-}\mathcal{L})} \leq g_N^{(\sigma)} \in [\Delta_N, 1)$. The quantities $g_N^{(\mathcal{L})}$ and $g_N^{(\sigma)}$ involve a maximum over all intervals (related to γ) because we estimate over different intervals where the volatility might be non-linear. So, on each interval, the result on the convergence speed requires a condition which allows to apply Lemma 5.2.

Theorem 4.5 when $\sigma_\star$ is known. When $\sigma_\star$ is known, there is no need to replace it by its estimator. Theorem 4.5 holds with weaker assumptions.
In Hypothesis $\mathcal{H}_{\mathcal{L}}$ the moment of order 4 can be replaced by the moment of order $(2q \max_{j=0,\cdots,d} \gamma_j)$ with $q > 1$.

In the precise case $\max_{j=0,\cdots,d}(\gamma_j) = 1/2$, the p in hypothesis $\mathcal{H}_{q\text{-}\mathcal{L}}$, can be chosen such that $2 < p < 2a_-/\sigma_-^2$ (see Proposition A.3). This condition is the one in [5, Proposition 5] in the case of the CIR process (solution to (1.1) with $\gamma = 1/2$).
For the QMLE, we can improve the conditions in the case $\gamma_d \geq 1/2$, by considering a different discretization of the QMLE (4.3): take $\mathcal{M}^{\cdot,1}_{T_N,N}$ instead of $M^{\cdot,1}_{T_N,N}$. In this case, we have a different estimator, and assumption $\mathcal{H}_{q\text{-}\mathcal{L}}$ can be replaced by the weaker finite $2\max_{j=0,\cdots,d}\gamma_j$-nd moment of μ. Furthermore, the additional condition appearing in the asymptotic normality property can be replaced by the weaker condition as well, namely $\lim_{N\to+\infty} T_N g_N^{(\sigma)} = 0$.

6 Numerical Experiments

In this section, we implement the MLE and QMLE based on discrete observations on simulated and US interest rates data.

6.1 Simulated Data

In this section, we investigate the efficiency of our estimators on simulated data. We simulate the T-CKLS process combining known Euler-Type schemes on different intervals such as the implicit scheme in [1, equation (3)] when the process is a CIR or a drifted version of the scheme in [32] when the diffusion coefficient is non-linear. More precisely we use the following scheme. Given $X_0 \in (0,\infty)$ and $(G_k)_{k\in\mathbb{N}}$ a sequence of i.i.d. standard Gaussian random variables, next we set $X_0^{(n)} = X_0$ and , we define for all $k \in \mathbb{N}$

$$X^{(n)}_{(k+1)/n} := \left| X^{(n)}_{k/n} + \frac{1}{n}\left(a(X^{(n)}_{k/n}) - b(X^{(n)}_{k/n})X^{(n)}_{k/n}\right) + \frac{1}{\sqrt{n}}\sigma(X^{(n)}_{k/n})(X^{(n)}_{k/n})^{\gamma(X^{(n)}_{k/n})}G_k \right|.$$

Further discussion about the most suitable numerical scheme for T-CKLS is beyond the purpose of this document. To estimate the parameters from the simulated data, we use the estimators from discrete observations in Section 4.1. The implementation has been done using Matlab and parameters are as in Table 1.

a_0	b_0	σ_0	γ_0
0.3	0.2	0.2	0.5

a_1	b_1	σ_1	γ_1
0	0	0.4	0

a_2	b_2	σ_2	γ_2
0.3	0.2	0.2	0.5

r_1	r_2
1	1.5

Table 1: Simulations parameters.

Firstly, we illustrate Theorem 4.5 for the drift parameters $\theta_\star = (a_i, b_i)_{i=0}^{d=2}$, diffusion parameters $\sigma_\star$, γ and two thresholds r_1, r_2 given in Table 1. We consider a process which follows a CIR dynamic close to 0 and far away from 0 and is a BM on an intermediate bounded interval. We simulate 10^4 trajectories of the T-CKLS with two threshold the set of numerical parameters $(T, N) = (10^3, 10^6)$ with starting condition determined as follows. As the process is supposed to be stationary, we first simulate one trajectory starting from

$X_0 > 0$ chosen arbitrarily, say $X_0 = 1$, with the scheme $X^{(n)}$ given above with $n = kN/T$ for some $k \in \mathbb{N} \setminus \{0\}$ (we took $k = 1$). Then we consider the final value of the latter trajectory as initial condition of the 10^4 trajectories.

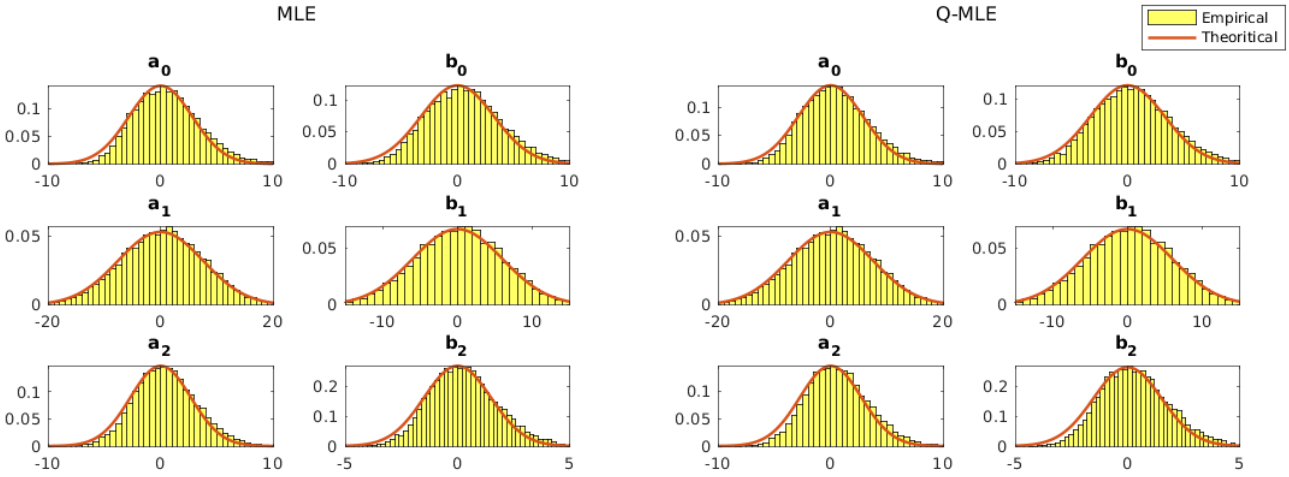

Figure 1: Asymptotic normality property in Theorem 4.5, with parameters as in Table 1. We plot the theoritical distribution using (7.1) and compare with the empirical distribution on 10^4 trajectories.

Remark that, despite the set of numerical parameters (T, N) does not satisfy the conditions for the asymptotic normality in Theorem 4.5, numerics shows good results.
We compare the estimators by means of relative root-mean-square error (RMSE) and mean-error (ME) in Table 2.

Estimator	Relative RMSE	ME
$(a^{0,(\mathcal{L})}_{T_N,N}, a^{1,(\mathcal{L})}_{T_N,N}, a^{2,(\mathcal{L})}_{T_N,N})$	(0.3206, 0.2341, 0.2921)	(0.0154, 0.0092, 0.0156)
$(b^{0,(\mathcal{L})}_{T_N,N}, b^{1,(\mathcal{L})}_{T_N,N}, b^{2,(\mathcal{L})}_{T_N,N})$	(0.5500, 0.1857, 0.2439)	(0.0173, 0.0069, 0.0093)
$(a^{0,(q\text{-}\mathcal{L})}_{T_N,N}, a^{1,(q\text{-}\mathcal{L})}_{T_N,N}, a^{2,(q\text{-}\mathcal{L})}_{T_N,N})$	(0.3233, 0.2341, 0.2965)	(0.0170, 0.0092, 0.0191)
$(b^{0,(q\text{-}\mathcal{L})}_{T_N,N}, b^{1,(q\text{-}\mathcal{L})}_{T_N,N}, b^{2,(q\text{-}\mathcal{L})}_{T_N,N})$	(0.5555,0.1857, 0.2478)	(0.0182,0.0069, 0.0112)
$(\sigma^0_{T_N,N}, \sigma^1_{T_N,N}, \sigma^2_{T_N,N})$	(0.0088, 0.0087, 0.0015)	(0.0087, -0.0034 , 0.001)

Table 2: Table of the relative RMSE and ME for the estimator of $(\theta_\star, \sigma_\star)$ using the MLE (4.8), QMLE (4.9) and volatility estimator (4.10).

Observe that the MLE gives a better estimation of the drift parameters, which is easily explained by the fact that the likelihood contains more information about the model. In general the MLE tends to have a better RMSE and ME than the QMLE. Applying the estimator on several data set, we remark that that the QMLE have a greater variance than the MLE.

Remark 6.1 (Threshold estimation)**.** *The thresholds can be estimated using the method proposed in [30] based on continuous observations. We use this procedure in the next*

section. It is a QMLE-based method without an explicit expression for the threshold estimator. Consequently, the numerical cost of this method increases with the number of thresholds. However, satisfactory results can still be achieved in the case of two thresholds. Mathematical study of the estimator for discrete observations should be conducted.

Remark 6.2 (Diffusion exponent estimation)**.** *The parameter γ was estimated in [26] for non-threshold CKLS. A study on the estimator properties for discrete observations in the context of CKLS or its generalisation to T-CKLS in necessary. However we expect that the estimator converges quite slowly especially around 0.*

6.2 Interest rates analysis

In this section, we apply our estimators to the ten-years US treasury rate based on the Federal Reserve Bank's H15 data set. We apply a discretized version of the test to evaluate the existence of one threshold (see [30], see Remark 6.1) or more thresholds (see [31]) in this dataset. In [31], the authors introduce a test statistic to identify thresholds in the drift term of a diffusion model: detect their presence and estimate them. They also develop a computationally efficient approach to calibrate the p-value and extend the test to detect multiple thresholds. Inspired by [33], we exploit here a discretized version of the method presented in [31] and combine with the drift MLE and $\sigma_\star$ estimator considered in this document.

General step. Let us describe the test to be applied at a general step and then the procedure describing the steps. Suppose that there are m thresholds, and that we look for the presence of an additional threshold on the k-th interval $I_k = (r_{k-1}, r_k)$ (we know the value of r_{k-1}, r_k: either known or estimated in previous steps). We consider the hypothesis:

$$\begin{cases} H_0: \text{ Null hypothesis} & m \text{ thresholds} \\ H_1: \text{ Alternative hypothesis} & (m+1) \text{ thresholds} \end{cases} \tag{6.1}$$

Under the null hypothesis H_0, the model has m thresholds. Under hypothesis H_1, there is an additional threshold $\bar{r}$ in the k-th regime meaning that the sequence of thresholds becomes $-\infty = r_0 < r_1 < \cdots < r_{k-1} < \bar{r} < r_k < \cdots < r_m$. The quasi-likelihood ratio test statistic is given by

$$T = \sup_{\bar{r} \in [a,b]} T(\bar{r})$$

with

$$T(\bar{r}) := 2\left(q\text{-}\mathcal{L}_{T_N,N}\left(\theta^{(H_1)}_{T_N,N}(m+1,\bar{r})\right) - q\text{-}\mathcal{L}_{T_N,N}\left(\theta^{(H_0)}_{T_N,N}(m)\right)\right) \tag{6.2}$$

where a and b are 20 and 80 percentiles of the data in $[r_{k-1}, r_k]$, the value $\theta^{(H_1)}_{T_N,N}(m+1,\bar{r})$ is the drift MLE under hypothesis H_1 with additional threshold given by $\bar{r}$, and $\theta^{(H_0)}_{T_N,N}(m)$ is the drift MLE of the model under hypothesis H_0. The MLE is given in Proposition 4.1, with volatility parameter $\sigma_\star$ estimated by (4.10).
We compute the statistics for some values of $\bar{r}$, say $\bar{r}_j := a(1-j/n)+bj/n$, $j \in \{0,1,\ldots,n\}$ (we choose $n = 10^3$). Then we take as an estimator for the threshold r_j the $\bar{r}_j$ which maximises $T(\bar{r})$ and the observed test statistics T_{data} is then the quantity $T(\bar{r}_j)$.
The distribution of the test statistic (6.2) is obtained using a bootstrap method. To compute the p-value we simulate 10^3 trajectories of the process with the parameters

under H_0, we compute T_j the statistics on the j-th trajectory. Then the p-value is given by $\#\{j\colon T_{data} < T_j\}/10^3$.
We fix the significance level at the conventional 5%.

Procedure. We apply the test above as follows, in a sequential procedure. We first test for the presence of a threshold on the data: applying this test for $m = 0$. If the test is significant, then we take as estimator for the threshold $\widehat{r}_1$ the $\bar{r}$ which realises the maximum in the test statistics. This threshold divides the state space into two intervals. We then we test the presence of thresholds on each of the two intervals, starting from the left to the right. On each interval we consider, if the test is significant, we keep dividing the interval into two sub-intervals and so on. Once we do not have evidence of new thresholds in the interval we are considering, we go to the next interval.

Application to ten years US treasury rate. We consider the ten-years US Treasury rate and we adopt the convention that the daily time intervals is $\mathrm{d}t = 0.046$, where one unit of time represents one month. We assume that the data follow a T-CIR dynamics, *i.e.* $\gamma \equiv 1/2$.
We consider the ten-years US Treasury rate for two different time windows: Jan 2016 - Dec 2019, and Jan 2020 - Jan 2024 represented in Figure 2.

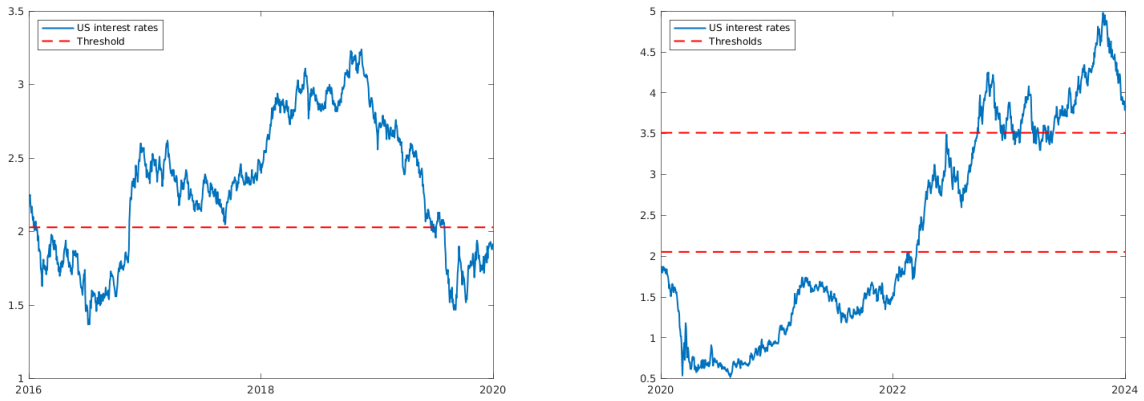

Figure 2: The figure shows the interest rate daily data (solid line) for the time windows Jan 2016 - Dec 2019, and Jan 2020 - Jan 2024. The fitted thresholds are represented by the dashed lines.

Let us consider the time window Jan 2016 - Dec 2019. The threshold test (6.1) for $m = 0$ is significant and the threshold estimation is $r_1 = 2.0303$. We apply the test (6.1) to detect a threshold on $(0, r_1)$. It is not significant. The same conclusion holds testing for threshold presence on $(r_1, +\infty)$.
On the time window Jan 2020 - Jan 2024 the threshold test (6.1) for $m = 0$ is significant and we estimate the threshold $r_1 = 2.0507$. There is no evidence of additional thresholds on $(0, r_1)$, and the null hypothesis is rejected for the existence of a threshold $r_2 = 3.5112$ on $(r_1, +\infty)$. Instead, the tests 6.1 with $m = 2$ to find a threshold on (r_1, r_2) and $(r_2, +\infty)$, the null hypothesis H_0 is not-rejected.

Jan 2016 - Dec 2019		Jan 2020 - Jan 2024	
Estimator	Value	Estimator	Value
$(a^{0,(\mathcal{L})}_{T_N,N}, a^{,(\mathcal{L})}_{T_N,N})$	(1.6434 , 0.1713)	$(a^{0,(\mathcal{L})}_{T_N,N}, a^{1,(\mathcal{L})}_{T_N,N}, a^{2,(\mathcal{L})}_{T_N,N})$	(0.2013, 0.5826, -0.0207)
$(b^{0,(\mathcal{L})}_{T_N,N}, b^{1,(\mathcal{L})}_{T_N,N})$	(0.9410, 0.0723)	$(b^{0,(\mathcal{L})}_{T_N,N}, b^{1,(\mathcal{L})}_{T_N,N}, b^{2,(\mathcal{L})}_{T_N,N})$	(0.1556, 0.0670, 0.0236)
$(\sigma^0_{T_N,N}, \sigma^1_{T_N,N})$	(0.1616, 0.1053)	$(\sigma^0_{T_N,N}, \sigma^1_{T_N,N}, \sigma^2_{T_N,N})$	(0.2129, 0.2091, 0.1807)
$r^1_{T_N,N}$	(2.0303)	$(r^1_{T_N,N}, r^2_{T_N,N})$	(2.0507, 3.5112)

Table 3: Estimated parameters corresponding to Figure 2.

Therefore, we conclude that there is a single threshold in the time window Jan 2016 - Dec 2019 and two thresholds in Jan 2020 - Jan 2024. In Table 3, we summarize the values obtained for each of the fitted parameters using the estimators.

7 Proofs

In this section, recalling some well known results, we show under which conditions on the parameters the process admits a stationary distribution. Then, we prove Theorem 4.3, Theorem 4.5, Lemma 5.2, Proposition 5.3 and Proposition 5.4.

Notation. We sometimes denote a_0, b_0, σ_0, γ_0 by a_-, b_-, σ_-, γ_- and a_d, b_d, σ_d, γ_d by a_+, b_+, σ_+, γ_+. These parameters determine the behavior of the process and moment properties. We use the alternative notation, when we wish to emphasize their role.

7.1 The process: properties of solutions and conditions for the stationary distribution

T-CLKS shows several behaviors, it may behave as an OU process on some intervals, a CIR or CKLS in others. Let us recall that we assumed that our process spends 0 time at 0. The state space of T-CKLS is determined by the process behavior around 0, in particular at I_0 where it behaves as a standard CKLS process. The regime of the process (transient, recurrent, positive-recurrent) is also determined by the behavior at I_0 and I_d. Let us introduce the scale function S and the speed measure $m(x)\,\mathrm{d}x$. The interested reader could refer to [6, II.4] for a summary or find more details e.g. in [29, Chapter VII, Section 3]. The scale function is continuous, unique up to a multiplicative constant, and its derivative satisfies $S'(x) = \exp\left(\int_{r_1}^x \frac{2(a(y)-b(y)y)}{\sigma(y)^2 y^{2\gamma(y)}}\,\mathrm{d}y\right)$. The speed measure is given by $m(x)\,\mathrm{d}x = \frac{2}{\sigma(x)|x|^{\gamma(x)}S'(x)}\,\mathrm{d}x$. The state of space of the T-CKLS process, denoted $I = \cup_{j=0}^d I_j$, depends on the value of the parameters in I_0 (a_-, σ_- and γ_-). The Feller boundary classification criteria (see e.g. [29]) implies the following lemma.

Lemma 7.1. *Let X be the solution to the SDE (2.1).*

- *If $\gamma_- = 0$, the state of space of the process is $I = \mathbb{R}$.*
- *The state space is $I = [0, +\infty)$ and the point 0 is instantaneously reflecting, if $0 < \gamma_- < 1/2$ or if $\gamma_- = 1/2$ and $a_- \leq \sigma_-^2/2$.*
- *If $\gamma_- \in [1/2, 1]$, then $I = (0, +\infty)$ and 0 is an unattainable boundary.*

The regime of the process can be obtained by properties on the scale function and the spead measure (see e.g. [29, Exercice 3.15 in Chapter X] and [16, Theorem 20.15]). We recall that, in the recurrent case the measure $m(x)\,\mathrm{d}x$ is a stationary measure and in the positive recurrent (we say *ergodic*) case it is a finite measure, so it can be renormalized to the stationary distribution:

$$\mu(\mathrm{d}x) = \frac{m(x)}{\int_I m(y)\,\mathrm{d}y}\,\mathrm{d}x. \tag{7.1}$$

The recurrent positivity property of the process only depends on the parameters below the first threshold (on I_0) and above the last threshold (on I_d). In the following table, we give conditions on the parameters $a_\pm$, $b_\pm$ and $\sigma_\pm$ depending on the value of $\gamma_\pm$ such that the process is ergodic (admits a stationary distribution).

	$a_- > 0$ and $b_- \in \mathbb{R}$, or
$\gamma_- = 1$	$a_- = 0$ and $b_- \leq -\frac{\sigma_-^2}{2}$
$\gamma_- \in (0,1)$	no other choice
$\gamma_- = 0$	$a_- = 0$ and $b_- > 0$

	$a_+ \in \mathbb{R}$ and $b_+ > 0$, or
$\gamma_+ = 1$	$a_+ \in \mathbb{R}$ and $b_+ \in \big[-\frac{\sigma_+^2}{2}, 0\big]$
$\gamma_+ \in (1/2,1)$	$a_+ \in \mathbb{R}$ and $b_+ = 0$
$\gamma_+ \in [0,1/2]$	$a_+ < 0$ and $b_+ = 0$

Table 4: Parameter conditions for ergodicity of T-CKLS process X solution to (2.1).

7.2 Proof of Theorem 4.3

This proof relies on Lemma 5.2 whose proof is provided in Section 7.4.
We study the asymptotic behavior of $(\sigma_{T_N,N})^2 - \sigma_\star^2$. For $j \in \{0,\cdots,d\}$, by Proposition 3.4, on the event $\{Q_T^{j,0} > 0\}$, we have:

$$(\sigma^j_{T_N,N})^2 - \sigma_j^2 = \sigma_j^2 \frac{Q^{j,2\gamma_j}_{T_N} - Q^{j,2\gamma_j}_{T_N,N}}{Q^{j,2\gamma_j}_{T_N,N}} + \frac{\mathbf{Q}^j_{T_N,N} - \mathbf{Q}^j_{T_N}}{Q^{j,2\gamma_j}_{T_N,N}}.$$

By equations (3.8)-(3.7), we have:

$$\left|\mathbf{Q}^j_{T_N} - \mathbf{Q}^j_{T_N,N}\right| \leq 2\left|M^{j,1}_{T_N} - M^{j,1}_{T_N,N}\right| + 2\max(|r_j|,|r_{j+1}|)\sum_{i=j}^{d}\left|M^{i,0}_{T_N} - M^{i,0}_{T_N,N}\right|,$$

for $j \in \{1,\cdots,d-1\}$ and for $j \in \{0,d\}$

$$\left|\mathbf{Q}^j_{T_N} - \mathbf{Q}^j_{T_N,N}\right| \leq 2\left|M^{j,1}_{T_N} - M^{j,1}_{T_N,N}\right| + 2|r_{(j+1)\wedge d}|\left|M^{j,0}_{T_N} - M^{j,0}_{T_N,N}\right|.$$

Then, we conclude by making use of item (1) in Lemma 5.2 and the fact that $\mathbb{P}\left(\lim_{T\to+\infty} Q_T^{j,0} > 0\right) = 1$.

7.3 Proof of Theorem 4.5

This proof relies on Lemma 5.2 that we prove in the next section. For all $N \in \mathbb{N}$ and $\ell \in \{\mathcal{L}, q\text{-}\mathcal{L}\}$ it holds that

$$\left(\theta^{(\ell)}_{T_N,N} - \theta_\star\right) = \left(\theta^{(\ell)}_{T_N,N} - \theta^{(\ell)}_{T_N}\right) + \left(\theta^{(\ell)}_{T_N} - \theta_\star\right).$$

The second term on the right hand side of the equality provides the asymptotic behavior by applying Theorem 3.5.
In the case $\ell = \mathcal{L}$, for $j \in \{0, \cdots, d\}$, using equations (4.8) and (3.4), each component of the first term can be rewritten as follows:

$$\left(\frac{Q^{j,k}_{T_N,N}}{Q^{j,-2\gamma_j}_{T_N,N} Q^{j,2-2\gamma_j}_{T_N,N} - (Q^{j,1-2\gamma_j}_{T_N,N})^2} - \frac{Q^{j,k}_{T_N}}{Q^{j,-2\gamma_j}_{T_N} Q^{j,2-2\gamma_j}_{T_N} - (Q^{j,1-2\gamma_j}_{T_N})^2} \right) \mathcal{M}^{j,m}_{T_N} + \frac{Q^{j,k}_{T_N,N}(\mathcal{M}^{j,m}_{T_N,N} - \mathcal{M}^{j,m}_{T_N})}{Q^{j,-2\gamma_j}_{T_N,N} Q^{j,2-2\gamma_j}_{T_N,N} - (Q^{j,1-2\gamma_j}_{T_N,N})^2},$$

with $k \in \{-2\gamma_j, 1-2\gamma_j, 2-2\gamma_j\}$ and $m \in \{-2\gamma_j, 1-2\gamma_j\}$. Let $\bar{r} := \max\{|r_j|, j = 1, \ldots, d\}$. Then, using formula (5.1) and (5.2), for $j \in \{1, \cdots, d-1\}$, we have

$$|\mathcal{M}^{j,m}_{T_N} - \mathcal{M}^{j,m}_{T_N,N}| \leq \bar{r}^m \sum_{i=j}^{d} |M^{i,0}_{T_N} - M^{i,0}_{T_N,N}| + \frac{|m|}{2} \left| \left(\sigma^j_{T_N,N}\right)^2 - \sigma_j^2 \right| \left| Q^{j,m+2\gamma_j-1}_{T_N} - Q^{j,m+2\gamma_j-1}_{T_N,N} \right|.$$

and for $j \in \{0, d\}$:

$$|\mathcal{M}^{j,m}_{T_N} - \mathcal{M}^{j,m}_{T_N,N}| \leq \bar{r}^m |M^{j,0}_{T_N} - M^{j,0}_{T_N,N}| + \frac{|m|}{2} \left| \left(\sigma^j_{T_N,N}\right)^2 - \sigma_j^2 \right| \left| Q^{j,m+2\gamma_j-1}_{T_N} - Q^{j,m+2\gamma_j-1}_{T_N,N} \right|.$$

This, Item (2) in Lemma 5.2 and Theorem 4.3 ensure that

$$T_N^{-1/\lambda} \left| Q^{j,k}_{T_N,N} - Q^{j,k}_{T_N} \right| \xrightarrow[N\to+\infty]{\mathbb{P}} 0 \quad \text{and} \quad T_N^{-1/\lambda} \left| \mathcal{M}^{j,m}_{T_N,N} - \mathcal{M}^{j,m}_{T_N} \right| \xrightarrow[N\to+\infty]{\mathbb{P}} 0$$

with $\lambda = 1$ to get the consistency of the MLE and $\lambda = 2$ for the speed of convergence.
The case $\ell = q\text{-}\mathcal{L}$, the first term $\left(\theta^{(\ell)}_{T_N,N} - \theta^{(\ell)}_{T_N}\right)$ works analogously, exploiting Item (3) in Lemma 5.2.

7.4 Proof of Lemma 5.2

This proof relies on Propositions 5.3 and 5.4, that we prove in the next sections. Hypothesis $\mathcal{H}_{\mathcal{L}}$ and $\mathcal{H}_{q\text{-}\mathcal{L}}$ ensure that we can apply these propositions.
In this proof we use the round ground notation $\lfloor t \rfloor_{\Delta_N} := t_k$ for $t \in [t_k, t_{k+1}) \subseteq [t_k, t_k + \Delta_N]$. Moreover, without loss of generality, we assume $T_N \leq N$ for all $N \in \mathbb{N}$.
We need to prove that $\mathbb{E}\left[|Q^{j,k}_{T_N} - Q^{j,k}_{T_N,N}|\right]$ and $\mathbb{E}\left[|M^{j,m}_{T_N} - M^{j,m}_{T_N,N}|\right]$ are $o(T_N^{1/\lambda})$. We follow a similar reasoning as in [24] to show that the proof is based on Proposition 5.3 and 5.4: For $j \in \{0, \cdots, d\}$ and $k \in \{-2\gamma_j, 1-2\gamma_j, 2-2\gamma_j, 2\gamma_j\} \cup \{-1, 0, 1, 2\}$:

$$\begin{aligned} Q^{j,k}_{T_N} - Q^{j,k}_{T_N,N} &= \int_0^{T_N} (X^k_t - X^k_{\lfloor t \rfloor_{\Delta_N}}) \mathbb{1}_{I_j}(X_t)\, \mathrm{d}t + \int_0^{T_N} X^k_{\lfloor t \rfloor_{\Delta_N}} (\mathbb{1}_{I_j}(X_t) - \mathbb{1}_{I_j}(X_{\lfloor t \rfloor_{\Delta_N}}))\, \mathrm{d}t \\ &= \int_0^{T_N} X^k_{\lfloor t \rfloor_{\Delta_N}} (\mathbb{1}_{X_t \in I_j, X_{\lfloor t \rfloor_{\Delta_N}} \notin I_j} - \mathbb{1}_{X_t \notin I_j, X_{\lfloor t \rfloor_{\Delta_N}} \in I_j})\, \mathrm{d}t + \int_0^{T_N} (X^k_t - X^k_{\lfloor t \rfloor_{\Delta_N}}) \mathbb{1}_{I_j}(X_t)\, \mathrm{d}t, \end{aligned}$$

therefore

$$\begin{aligned} \mathbb{E}\left[|Q^{j,k}_{T_N} - Q^{j,k}_{T_N,N}|\right] \leq \int_0^{T_N} \mathbb{E}\left[|X^k_t - X^k_{\lfloor t \rfloor_{\Delta_N}}|\right] \mathrm{d}t + \int_0^{T_N} \mathbb{E}\left[|X_{\lfloor t \rfloor_{\Delta_N}}|^k \mathbb{1}_{X_t \in I_j, X_{\lfloor t \rfloor_{\Delta_N}} \notin I_j}\right] \mathrm{d}t \\ + \int_0^{T_N} \mathbb{E}\left[|X_{\lfloor t \rfloor_{\Delta_N}}|^k \mathbb{1}_{X_t \notin I_j, X_{\lfloor t \rfloor_{\Delta_N}} \in I_j}\right] \mathrm{d}t. \end{aligned} \tag{7.2}$$

In a similar way, for $m \in \{0,1\}$ it holds

$$M^{j,m}_{T_N,N} - M^{j,m}_{T_N} = \int_0^{T_N} \left(X_t^m \mathbb{1}_{I_j}(X_t) - X^m_{\lfloor t \rfloor_{\Delta_N}} \mathbb{1}_{I_j}(X_{\lfloor t \rfloor_{\Delta_N}}) \right) (a(X_t) - b(X_t)X_t)\, \mathrm{d}t + \int_0^{T_N} \left(X_t^m \mathbb{1}_{I_j}(X_t) - X^m_{\lfloor t \rfloor_{\Delta_N}} \mathbb{1}_{I_j}(X_{\lfloor t \rfloor_{\Delta_N}}) \right) \sigma(X_t)(X_t)^{\gamma(X_t)}\, \mathrm{d}B_t.$$

Using Triangular inequality, Hölder's inequality, and Itô-isometry, we obtain

$$\begin{aligned}
&\mathbb{E}\left[|M^{j,m}_{T_N} - M^{j,m}_{T_N,N}|\right] \leq \int_0^{T_N} \mathbb{E}\left[|X_t^m - X^m_{\lfloor t \rfloor_{\Delta_N}}|(|a_j| + b_j|X_{\lfloor t \rfloor_{\Delta_N}}| + b_j|X_t - X_{\lfloor t \rfloor_{\Delta_N}})|)\right] \mathrm{d}t \\
&+ \int_0^{T_N} \mathbb{E}\left[|X_{\lfloor t \rfloor_{\Delta_N}}|^m \mathbb{1}_{X_t \in I_j, X_{\lfloor t \rfloor_{\Delta_N}} \notin I_j} \left(\max_{i=0,\cdots,d} |a_i| + \max_{i=0,\cdots,d} |b_i|(|X_{\lfloor t \rfloor_{\Delta_N}}| + |X_t - X_{\lfloor t \rfloor_{\Delta_N}}|) \right) \right] \mathrm{d}t \\
&+ \int_0^{T_N} \mathbb{E}\left[|X_{\lfloor t \rfloor_{\Delta_N}}|^m \mathbb{1}_{X_t \notin I_j, X_{\lfloor t \rfloor_{\Delta_N}} \in I_j} \left(\max_{i=0,\cdots,d} |a_i| + \max_{i=0,\cdots,d} |b_i|(|X_{\lfloor t \rfloor_{\Delta_N}}| + |X_t - X_{\lfloor t \rfloor_{\Delta_N}}|) \right) \right] \mathrm{d}t \\
&+ \sqrt{2} \max_{i=0,\cdots,d}(\sigma_i) \left(\int_0^{T_N} \mathbb{E}\left[(X_t^m - X^m_{\lfloor t \rfloor_{\Delta_N}})^2 (X_t)^{2\gamma(X_t)} + X^{2m}_{\lfloor t \rfloor_{\Delta_N}} (X_t)^{2\gamma(X_t)} \mathbb{1}_{X_t \in I_j, X_{\lfloor t \rfloor_{\Delta_N}} \notin I_j} \right. \right. \\
&\left. \left. + X^{2m}_{\lfloor t \rfloor_{\Delta_N}} (X_t)^{2\gamma(X_t)} \mathbb{1}_{X_t \notin I_j, X_{\lfloor t \rfloor_{\Delta_N}} \in I_j} \right] \mathrm{d}t \right)^{1/2}. \qquad (7.3)
\end{aligned}$$

Hence, for $\lambda \in \{1,2\}$, the proof of Proposition 5.2, reduces to prove the following statements:

$$\int_0^{T_N} \mathbb{E}\left[|X_t - X_{\lfloor t \rfloor_{\Delta_N}}|^p \right] \mathrm{d}t \quad \text{is} \quad o(T_N^{1/\lambda}), \qquad (7.4)$$

and

$$\int_0^{T_N} \mathbb{E}\left[|X_{\lfloor t \rfloor_{\Delta_N}}|^q \mathbb{1}_{X_t \notin I_j, X_{\lfloor t \rfloor_{\Delta_N}} \in I_j} \right] \mathrm{d}t \quad \text{and} \quad \int_0^{T_N} \mathbb{E}\left[|X_{\lfloor t \rfloor_{\Delta_N}}|^q \mathbb{1}_{X_t \in I_j, X_{\lfloor t \rfloor_{\Delta_N}} \notin I_j} \right] \mathrm{d}t \quad (7.5)$$

are $o(T_N^{1/\lambda})$ where p and q depend on the value of m and k, the coefficients of $M^{j,m}$ and $Q^{j,k}$. In Table 5, we summarize the different values of p and q according to the estimators we study. Hypothesis $\mathcal{H}_{\mathcal{L}}$ and $\mathcal{H}_{q\text{-}\mathcal{L}}$ ensure that the moments corresponding to Table 5 are finite.

We are now reduced to prove equations (7.4) and (7.5) instead of (7.2) and (7.3). To do so, we exploited Hypotheses $\mathcal{H}_{\mathcal{L}}$ and $\mathcal{H}_{q\text{-}\mathcal{L}}$ on moments of μ. More precisely, in the case of the QMLE and the volatility estimator, the hypothesis $\mathcal{H}_{q\text{-}\mathcal{L}}$ was used in the bounds for (7.3) with $m = 1$. Instead, for the MLE, the hypothesis $\mathcal{H}_{\mathcal{L}}$ was used to bound (7.2) with $k = -2\gamma_j$ and $k = 2 - 2\gamma_j$.

$\theta^{(1)}_{T_N,N}$	$m \in \{0\}$ $k \in \{-2\gamma_j, -1, 0, 1-2\gamma_j, 2-2\gamma_j\}$	$p \in \{1, 2, 4\gamma_j\}$ $q \in \{-1, -2\gamma_j, 1-2\gamma_j, 0, 2-2\gamma_j\}$
$\theta^{(2)}_{T_N,N}$	$m \in \{0,1\}$ $k \in \{0,1,2\}$	$p \in \{1,2,4\}$ $q \in \{0,1,2,4\}$
$\sigma_{T_N,N}$	$m \in \{0,1\}$ $k \in \{2\gamma_j\}$	$p \in \{1,2\}$ $q \in \{0,1,2\}$

Table 5: Summary of quantity control in the MLE (4.8), QMLE (4.9) and the volatility estimator (4.10).

Note that, equations (7.5) can be simplified as follows. The tower property of conditional expectation and Markov property imply that

$$\mathbb{E}\left[|X_{\lfloor t\rfloor_{\Delta_N}}|^q \mathbb{1}_{X_t\notin I_j, X_{\lfloor t\rfloor_{\Delta_N}}\in I_j}\right] \leq \mathbb{E}\left[|X_{\lfloor t\rfloor_{\Delta_N}}|^q \mathbb{P}_{X_{\lfloor t\rfloor_{\Delta_N}}}\left(\tau_{I_j}^{\xi^j} < \Delta_N\right)\mathbb{1}_{I_j}(X_{\lfloor t\rfloor_{\Delta_N}})\right],$$

and

$$\mathbb{E}\left[|X_{\lfloor t\rfloor_{\Delta_N}}|^q \mathbb{1}_{X_t\in I_j, X_{\lfloor t\rfloor_{\Delta_N}}\notin I_j}\right] \leq \sum_{\substack{i=0\\ i\neq j}}^{d} \mathbb{E}\left[|X_{\lfloor t\rfloor_{\Delta_N}}|^q \mathbb{P}_{X_{\lfloor t\rfloor_{\Delta_N}}}\left(\tau_{I_i}^{\xi^i} < \Delta_N\right)\mathbb{1}_{I_i}(X_{\lfloor t\rfloor_{\Delta_N}})\right],$$

where ξ^j is a CKLS process with parameter $(a_j, b_j, \sigma_j, \gamma_j)$ starting at $\xi_0^j = X_{\lfloor t\rfloor_{\Delta_N}}$ but driven by a BM independent of $\mathcal{F}_{\lfloor t\rfloor_{\Delta_N}}$ and $\tau_{I_j}^{\xi_j}$ is the first hitting time of the interval I_j for the process ξ^j.
The fact that (7.5) and (7.4) hold, follows from Proposition 5.4 and Proposition 5.3.

7.5 Proof of Proposition 5.3

Given s and t such that $0 \leq s < t$, we show that for every $m \geq 1$ such that μ admits finite m-th moment, there exists a constant $C \in (0,\infty)$ depending only on m and the parameters of the process such that $\mathbb{E}\left[|X_t - X_s|^m\right] \leq C(t-s)^{m/2}$.
By the triangular inequality,

$$\begin{aligned}|X_t - X_s| &\leq \int_s^t |a(X_u) - b(X_u)X_u|\,\mathrm{d}u + \left|\int_s^t \sigma(X_u)(X_u)^{\gamma(X_u)}\,\mathrm{d}B_u\right| \\ &\leq (t-s)\max_{i=0,\cdots,d}|a_i| + \max_{i=0,\cdots,d}|b_i|\int_s^t |X_u|\,\mathrm{d}u + \left|\int_s^t \sigma(X_u)(X_u)^{\gamma(X_u)}\,\mathrm{d}B_u\right|.\end{aligned}$$

Then, Jensen's inequality ensures that for $m \geq 1$ it holds that

$$\begin{aligned}\mathbb{E}\left[|X_t - X_s|^m\right] \leq\ & 2^{2m-2}\max_{i=0,\cdots,d}|a_i|(t-s)^m + 2^{2m-2}\max_{i=0,\cdots,d}|b_i|(t-s)^{m-1}\int_s^t \mathbb{E}\left[|X_u|^m\right]\mathrm{d}u \\ &+ 2^{m-1}\mathbb{E}\left[\left|\int_s^t \sigma(X_u)(X_u)^{\gamma(X_u)}\,\mathrm{d}B_u\right|^m\right].\end{aligned}$$

Since X_0 is distributed as the stationary distribution μ and μ admits finite m-th moment, then $\sup_{u\in[s,t]}\mathbb{E}\left[|X_u|^m \mathbb{1}_{I_d}(X_u)\right] < \infty$.
Burkholder-Davis-Gundy inequality implies that

$$\mathbb{E}\left[\left|\int_s^t \sigma(X_u)(X_u)^{\gamma(X_u)}\,\mathrm{d}B_u\right|^m\right] \leq \mathbb{E}\left[\left(\int_s^t \max_{i=0,\cdots,d}|\sigma_i|(X_u)^{2\gamma(X_u)}\,\mathrm{d}u\right)^{m/2}\right].$$

Then, we distinguish the case $m \geq 2$ and $m \in [1,2)$. In both cases we apply Hölder's inequality but in a differnet way. If $m \geq 2$, we obtain

$$\begin{aligned}\mathbb{E}\left[\left|\int_s^t \sigma(X_u)(X_u)^{\gamma(X_u)}\,\mathrm{d}B_u\right|^m\right] &\leq (t-s)^{m/2-1}\left(\max_{i\in\{0,\cdots,d\}}|\sigma_i|\right)^{m/2}\int_s^t \mathbb{E}\left[|X_u|^{m\gamma_+}\right]\mathrm{d}u \\ &\leq C(t-s)^{m/2}, \qquad (7.6)\end{aligned}$$

since, $\sup_{u\in[s,t]}\mathbb{E}\left[|X_u|^{m\gamma_+}\mathbb{1}_{I_d}(X_u)\right]<\infty$. If $m\in[1,2)$, we reduce to the previous case ($m\geq 2$):

$$\mathbb{E}\left[\left|\int_s^t \sigma(X_u)(X_u)^{\gamma(X_u)}\,\mathrm{d}B_u\right|^m\right]\leq\left(\mathbb{E}\left[\left|\int_s^t \sigma(X_u)(X_u)^{\gamma(X_u)}\,\mathrm{d}B_u\right|^{2m}\right]\right)^{1/2}\leq C(t-s)^{m/2}.$$

The proof is thus completed.

7.6 Proof of the key result: Proposition 5.4

For all $j\in\{0,\cdots,d\}$, let ξ^j denote a standard CKLS process with parameters $(a_j,b_j,\sigma_j,\gamma_j)$ starting at X_s. Let $s,t\in[0,\infty)$ be fixed such that $0\leq s<t$. Note that:

$$\begin{aligned}&\mathbb{E}\left[|X_s|^m\mathbb{P}_{X_s}\left(\tau^{\xi^j}_{I_j}<t-s\right)\right]\\&\leq\mathbb{E}\left[|X_s|^m\left(\mathbb{P}_{X_s}\left(\tau^{\xi^j}_{r_j,\searrow}<t-s\right)+\mathbb{P}_{X_s}\left(\tau^{\xi^j}_{r_{j+1},\nearrow}<t-s\right)\right)\mathbb{1}_{I_j}(X_s)\right],\end{aligned}$$

where $\tau^{\xi^j}_{r_j,\searrow}$ is the first hitting time from above of the level r_j, $\tau^{\xi^j}_{r_{j+1},\nearrow}$ is the first hitting time from below of the level r_{j+1} of the process ξ^j.
Without loss of generality, we reduce to show that:

$$\mathbb{E}\left[|X_s|^m\mathbb{P}_{X_s}\left(\tau^{\xi^+}_{r_d,\searrow}<t-s\right)\mathbb{1}_{I_d}(X_s)\right]\leq C_2(t-s)^{1/2},\tag{7.7}$$

and

$$\mathbb{E}\left[|X_s|^m\mathbb{P}_{X_s}\left(\tau^{\xi^-}_{r_1,\nearrow}<t-s\right)\mathbb{1}_{I_0}(X_s)\right]\leq C_1(t-s)^{1/2},\tag{7.8}$$

where C_1 and C_2 are strictly positive constant.
Indeed, in the other cases, X_s belongs in I_j for $j\notin\{0,d\}$, which is compact, and the desired inequality can be deduced using a similar reasoning.

7.6.1 Bounds on the first hitting time from above of the level r_d: (7.7)

Let $\xi^+:=\xi^d$, we focus on the case $\gamma_+\in[0,1)$. The case $\gamma_+=1$ can be proven using a similar reasoning, the proof is thus omitted. Let us recall that the parameters $(a_+,b_+,\sigma_+,\gamma_+)$ satisfy the ergodicity conditions in Table 4 in Section 7.1, in particular $b_+\geq 0$.
The main idea of this proof is to bound the first hitting time by the hitting times of some drifted Brownian motions. To do so, we apply the Lamperti transform and we bound the process, over a well chosen time interval.
We define the process $(Y_u)_{u\geq 0}$ as follows. For all $u\geq 0$ let $Y_u=\psi(\xi^+_u)$ where $\psi(x)=\int_0^x\frac{1}{\sigma_+y^{\gamma_+}}\,\mathrm{d}y=\frac{x^{1-\gamma_+}}{\sigma_+(1-\gamma_+)}$ (Lamperti tranform). We denote ψ^{-1} as the inverse function of ψ, then Y is solution to the following SDE:

$$\mathrm{d}Y_u=\frac{a_+}{\sigma_+}(\psi^{-1}(Y_u))^{-\gamma_+}-b_+(1-\gamma_+)Y_u-\frac{1}{2}\sigma_+\gamma_+(\psi^{-1}(Y_u))^{\gamma_+-1}\,\mathrm{d}u+\mathrm{d}B_u.$$

Let $\varepsilon>0$ be fixed and let $\tau^Y_{[\psi(r_d),\psi(X_s)+\varepsilon]}$ denote the first hitting time of the boundary of $[\psi(r_d),\psi(X_s)+\varepsilon]$ of the process Y. Then, we have

$$\mathbb{P}_{X_s}\left(\tau^{\xi^+}_{r_d,\searrow}<t-s\right)=\mathbb{P}_{\psi(X_s)}\left(\tau^Y_{\psi(r_d),\searrow}<t-s\right)\leq\mathbb{P}_{\psi(X_s)}\left(\tau^Y_{[\psi(r_d),\psi(X_s)+\varepsilon]}<t-s\right).$$

By applying the Comparison Theorem (e.g. [13, p352]) until time $\tau^Y_{[\psi(r_d),\psi(X_s)+\varepsilon]}$, $Y-\psi(X_s)$ is bounded from above by a drifted BM B^{μ_+} starting from 0 and from below by a drifted BM B^{μ_-} starting from 0 with parameters:

$$\begin{cases} \mu_+ = \frac{|a_+|}{\sigma_+} r_d^{-\gamma_+} \\ \mu_- = -\frac{|a_+|}{\sigma_+} r_d^{-\gamma_+} - b_+(1-\gamma_+)(\psi(X_s)+\varepsilon) - \frac{1}{2}\sigma_+\gamma_+ r_d^{\gamma_+-1}. \end{cases}$$

Hence, the following inequality holds:

$$\mathbb{P}_{X_s}\left(\tau^{\xi^+}_{r_d,\searrow} < t-s\right) \leq \mathbb{P}_0\left(\tau^{B^{\mu_-}}_{\psi(r_d)-\psi(X_s),\searrow} < t-s\right) + \mathbb{P}_0\left(\tau^{B^{\mu_+}}_{\varepsilon,\nearrow} < t-s\right),$$

and classical results on the first hitting of a drifted Brownian motion (see [6]) yield

$$\mathbb{P}_0\left(\tau^{B^{\mu_+}}_{\varepsilon,\nearrow} < t-s\right) \leq K_1 e^{-\frac{\varepsilon^2}{2(t-s)}}, \tag{7.9}$$

and,

$$\mathbb{P}_0\left(\tau^{B^{\mu_-}}_{\psi(r_d)-\psi(X_s),\searrow} < t-s\right) \leq K_2 e^{-\frac{|\psi(r_d)-\psi(X_s)|^2}{2(t-s)} + b_+(1-\gamma_+)\psi(X_s)^2} f(X_s), \tag{7.10}$$

where K_1 and K_2 are two strictly positive constants, f is an explicit function which depends only on X_s and such that $\lim_{x\to+\infty} f(x)e^{-\psi(x)^2} = 0$.
Let us note that

$$\mu(x)\mathbb{1}_{x\geq r_d} = K_3 \frac{2}{\sigma_+^2 x^{1-2\frac{a_+}{\sigma_+^2}}} \exp\left(-b_+(1-\gamma_+)\psi(x)^2\right)\mathbb{1}_{x\geq r_d},$$

with K_3 a strictly positive constant. Since μ admits finite m-th moment by assumption, inequality (7.7) holds. We avoid details here, but the interested reader could appreciate the following remarks. When $t-s$ is close to 0, $e^{-K/(t-s)}$ decreases faster than any polynomial of $(t-s)$. So, up to splitting the integral to distinguish between when $\psi(x)-\psi(r_d)$ is small (e.g. $O(\sqrt{t-s})$) and when it isn't, we can easily compute the bounds. To avoid repetitions, we do not mention anymore this remark.

7.6.2 Bounds on the first hitting time from below of the level r_1: (7.8)

Let $\xi^- := \xi^0$, the parameters $(a_-, b_-, \sigma_-, \gamma_-)$ satisfy the conditions ensuring ergodicity in Table 4. We remark that, on a suitable time interval, the process $(\xi^-)^{2(1-\gamma_-)}$ can be bounded from above by the norm of a multi-dimensional Brownian motion. This leads to obtaining (7.8) under the assumption that $\gamma_- \in [0, 1/2]$. Instead, if $\gamma_- \in (1/2, 1]$, this bound is not enough. Thus, we additionally bound from below the Lamperti transform of the process ξ^- by a drifted Brownian motion.

The case $\gamma_- = 1/2$. We remind the following results, for $i \in \{1, \cdots, n\}$, we denote $\tilde{B} := (\tilde{B}^i)_{i\leq n}$ a n-dimensional Brownian motion. Itô formula and Levy characterization imply that

$$\forall u \geq 0, \quad \left|\left|\frac{\sigma_-^2}{4}\tilde{B}_u - \sqrt{\frac{X_s}{n}}\right|\right|_2^2 = Y_{u,n} = X_s + n\frac{\sigma_-^2}{4}u + \int_0^u \sigma_- \sqrt{Y_{v,n}}\,\mathrm{d}W_v.$$

Then $Y_{\cdot,n}$ is a CIR process whose coefficients satisfy the conditions for ergodicity in Table 4. Moreover we take n such that $0 < a_- < n\sigma_-^2/4$. So, by the Comparison Theorem, it holds a.s. that $\xi_u^- \leq Y_{u,n}$ for all $0 \leq u \leq \tau^{\xi^-}_{r_1,\nearrow}$ and then:

$$\mathbb{P}_{X_s}\left(\tau^{\xi^-}_{r_1,\nearrow} \leq t-s\right) \leq \mathbb{P}_{X_s}\left(\tau^{Y_{\cdot,n}}_{r_1,\nearrow} \leq t-s\right).$$

Moreover, we have:

$$\{\tau^{Y_{\cdot,n}}_{r_1,\nearrow} < t-s\} \subset \bigcap_{i=1}^{n}\left\{\forall u \leq t-s, -\sqrt{r_1} \leq \left(\frac{\sigma_-^2}{4}\tilde{B}^i_u - \sqrt{\frac{X_s}{n}}\right) \leq \sqrt{r_1}\right\}.$$

Then, by the Comparison Theorem and the symmetry of BM, we obtain the following inequality:

$$\mathbb{P}_{X_s}\left(\tau^{\xi^-}_{r_1,\nearrow} \leq t-s\right) \leq 4\mathbb{P}_0\left(\tau^{\tilde{B}^1}_{\sqrt{r_1}-\sqrt{X_s},\nearrow} \leq t-s\right) \leq K_4 e^{-\frac{\left(\sqrt{r_1}-\sqrt{X_s}\right)^2}{2(t-s)}}, \tag{7.11}$$

with K_4 a strictly positive constant. This, and the stationary distribution μ (7.1), if μ admits finite m-th moment, yield inequality (7.8).

The case $\gamma_- \in [0, 1/2)$. By applying Itô formula on $Y := (\xi^-)^{2(1-\gamma_-)}$, for all $u \geq 0$, we have:

$$\mathrm{d}Y_u = 2(1-\gamma_-)\left(a_- Y_u^{1-\frac{1}{2(1-\gamma_-)}} - b_- Y_u + \frac{\sigma_-^2(1-2\gamma_-)}{2}\right)\mathrm{d}u + 2(1-\gamma_-)\sigma_-\sqrt{Y_u}\,\mathrm{d}B_u$$

with $Y_0 = X_s^{2(1-\gamma_-)}$. Since $1-2\gamma_- > 0$, the Comparison Theorem ensures that for all $u \leq \tau^{Y}_{r^{2(1-\gamma_-)},\nearrow}$ the process Y is bounded from above by the norm of a n-dimensional Brownian motion. Hence, similarly to the case $\gamma_- = \frac{1}{2}$, we have:

$$\mathbb{P}_{X_s}\left(\tau^{\xi^-}_{r_1,\nearrow} \leq t-s\right) \leq 4\mathbb{P}_0\left(\tau^{\tilde{B}^1}_{r_1^{(1-\gamma_-)}-X_s^{(1-\gamma_-)},\nearrow} \leq t-s\right) \leq K_5 e^{-\frac{\left(r_1^{(1-\gamma_-)}-X_s^{(1-\gamma_-)}\right)^2}{2(t-s)}}, \tag{7.12}$$

with K_5 a strictly positive constant. We conclude analogously to the case $\gamma_- = \frac{1}{2}$.

The case $\gamma_- \in (1/2, 1)$. By Lamperti tranform, $Y_u := \mathrm{d}\psi(\xi_u^-)$ with $\psi(x) = \frac{x^{1-\gamma_-}}{\sigma_-(1-\gamma_-)}$ (in particular $Y_0 = \frac{X_s^{1-\gamma_-}}{\sigma_-(1-\gamma_-)}$). So,

$$\mathrm{d}Y_u = \left[\frac{a_-}{\sigma_-}\left((1-\gamma_-)Y_u\right)^{1-\frac{1}{(1-\gamma_-)}} - b_-(1-\gamma_-)Y_u - \frac{\gamma_-}{2(1-\gamma_-)}Y_u^{-1}\right]\mathrm{d}u + \mathrm{d}B_u.$$

By the Comparison Theorem, it holds a.s. for all $u \in [0, \leq \tau_{\left[\frac{\psi(X_s)}{2},\psi(r_1)\right]})$ that

$$B^{\mu_-}_u \leq Y_u \quad \text{and} \quad Y_u^2 \leq \left|\left|\frac{\sigma_-^2}{4}\tilde{B}_u - \frac{\psi(X_s)}{\sqrt{n}}\right|\right|_2^2,$$

where B^{μ_-} is drifted Brownian motion and $\tilde{B} := (\tilde{B}^i)_{i\leq n}$ a n-dimensional Brownian motion. Here the drift parameter μ_- and the dimension n are given by

$$\begin{cases} \mu_- = -\frac{|b_-|r_1^{1-\gamma_-}}{\sigma_-} - \frac{\gamma_-\sigma_-}{X_s^{1-\gamma_-}}, \\ n(X_s) = \left\lceil \frac{a_-}{\sigma_-^2(1-\gamma_-)}X_s^{1-2\gamma_-} + |b_-|(1-\gamma_-)\frac{r_1^{2(1-\gamma_-)}}{\sigma_-^2(1-\gamma_-)^2}\right\rceil. \end{cases}$$

Finally, we obtain the following inequality:

$$\begin{aligned}\mathbb{P}_{X_s}\left(\tau^{\xi^-}_{r_1,\nearrow} < t-s\right) &\leq \mathbb{P}_0\left(\tau^{B^{\mu_-}}_{-\frac{\psi(X_s)}{2},\searrow} < t-s\right) + 4\mathbb{P}_0\left(\tau^{\tilde{B}^1}_{\psi(r_1)^2-\psi(X_s)^2,\nearrow} < t-s\right) \\ &\leq K_6\left(e^{-\frac{|\psi(r_1)-\psi(X_s)|^2}{2(t-s)}} + e^{-\frac{\left(\psi(r_1)^2-\psi(X_s)^2\right)^2}{2(t-s)}}\right), \qquad (7.13)\end{aligned}$$

with K_6 a strictly positive constant.
Similarly to the previous cases, this, and the stationary distribution μ (7.1), if μ admits finite m-th moment, yield inequality (7.8).

The case $\gamma_- = 1$. One can conclude by bounding directly the Lamperti transform from above by a drifted Brownian motion as it has been done for the bounds on the first hitting time from above of the level r_d.

A Appendix: Auxiliary results

In this section, we provide some auxiliary results on well posedness of some integrals appearing in the likelihood in Proposition A.1 and on the finiteness of the moments of the stationary measure μ in Proposition A.3.
More precisely, we give some properties of the moments of the T-CKLS process. Some are straightforward applications of the ergodic properties.
The following proposition describes the behavior of various integrals of the T-CKLS process. It establishes whether the likelihood (2.4) is well defined or not.

Proposition A.1. *Let X solution to the SDE* (2.1).

1. *If $\gamma_- = 1/2$ and $a_- \geq \sigma_-^2/2$ or if $\gamma_- \in (1/2, 1] \cup \{0\}$, then*

$$\forall t \geq 0, \quad \int_0^t \frac{1}{X_s^{2\gamma_-}} \mathbb{1}_{I_0}(X_s)\,\mathrm{d}s < \infty \quad \mathbb{P}_{x_0} - \ a.s..$$

2. *If $\gamma_- = 1/2$ and $a_- < \sigma_-^2/2$ or if $\gamma_- \in [1/3, 1/2)$, then*

$$\forall t \geq 0, \quad \mathbb{P}_{x_0}\left(\int_0^t \frac{1}{X_s^{2\gamma_-}} \mathbb{1}_{I_0}(X_s)\,\mathrm{d}s = \infty\right) > 0.$$

Remark A.2 ($\gamma_- < 1/3$)**.** *Although we do not provide Proposition A.1 when $\gamma_- \in (0, 1/3)$, let us mention that, if X_0 is distributed according to the stationary measure μ and $\gamma_- \in (1/4, 1/3)$, then for all $t > 0$ it holds $\mathbb{E}_\mu\left[\int_0^t X_s^{-2\gamma_-} \mathbb{1}_{I_0}(X_s)\,\mathrm{d}s\right] < \infty$ and so for all $t > 0$ $\mathbb{P}_\mu\left(\int_0^t X_s^{-2\gamma_-} \mathbb{1}_{I_0}(X_s)\,\mathrm{d}s < \infty\right) = 1$. Instead, if $\gamma_- \in (0, 1/4]$, then for all $t > 0$ it holds $\mathbb{E}_\mu\left[\int_0^t X_s^{-2\gamma_-} \mathbb{1}_{I_0}(X_s)\,\mathrm{d}s\right] = \infty$.*

Proof. The first item follows from the fact that 0 is an unattainable boundary and continuity of the trajectories: the image of $[0, t]$ through each trajectory $s \mapsto X_s(\omega)$ is a compact of $]0, \infty[$.
The second item, for $\gamma_- = 1/2$ has been proven in [5] by using properties of the Laplace transform.

Let us consider the remaining case: $\gamma_- \in [1/3, 1/2)$. Since $1-\gamma_- \leq 2\gamma_-$, for every small $\varepsilon > 0$ we have

$$\mathbb{P}\left(\int_0^t \frac{1}{X_s^{1-\gamma_-}}\,\mathrm{d}s = \infty,\ \forall s \leq t\ X_s \leq \varepsilon\right) \leq \mathbb{P}\left(\int_0^t \frac{1}{X_s^{2\gamma_-}}\mathbb{1}_{I_0}(X_s)\,\mathrm{d}s = \infty\right).$$

Then by applying the Itô formula, and the Comparison Theorem, on the event $\{\forall s \leq t,\, X_s \leq \varepsilon\}$ it holds for all $s \leq t$ that $X_s^{2(1-\gamma_-)} \leq Y_s^\varepsilon$ where

$$\mathrm{d}Y_s^\varepsilon = 2(1-\gamma_-)\left(a_-\varepsilon^{1-\frac{1}{2(1-\gamma_-)}} - b_- Y_s^\varepsilon + \frac{\sigma_-^2(1-2\gamma_-)}{2}\right)\mathrm{d}s + 2(1-\gamma_-)\sigma_-\sqrt{Y_s^\varepsilon}\,\mathrm{d}B_s.$$

Let ε such that $\varepsilon < \min\left\{1, \left(\frac{\sigma_-^2\gamma_-}{2a_-}\right)^{1+1/(1-2\gamma_-)}\right\}$, then the process Y^ε is a CIR process and it reaches the point 0. Then, Theorem 2.3 in [27] ensures that

$$0 < \mathbb{P}\left(\int_0^t \frac{1}{\sqrt{Y_s^\varepsilon}}\,\mathrm{d}s = \infty\right) \leq \mathbb{P}\left(\int_0^t \frac{1}{X_s^{(1-\gamma_-)}}\,\mathrm{d}s = \infty,\quad \forall s \leq t\ X_s \leq \varepsilon\right).$$

This completes the proof. □

The following proposition describes the behaviour of the moments from the stationary distribution of the T-CKLS process in the ergodic regime. We recall that the stationary distribution μ is given by (7.1).

Proposition A.3. *Let $m \in [0, \infty)$ and assume that the conditions in Table 4 (ensuring that μ is the stationary distribution) hold. Then μ admits finite m-th moment unless it holds simultaneously $m \neq 0$, $b_+ = 0$, and $\gamma_+ \in [1/2, 1)$ in which case the m-th moment is finite if*

- *$\gamma_+ = 1/2$, $b_+ = 0$ and $a_+ < -m\frac{\sigma_+^2}{2}$.*
- *$\gamma_+ \in (1/2, 1)$, $b_+ = 0$ and $m < \gamma_+ - 1/2 < 1$.*
- *$\gamma_+ = 1$ and $b_+ \geq \frac{(m-1)\sigma_+^2}{2}$.*

The measure μ admits finite $-m$-th moment unless $\gamma_- \in (0, 1/2]$ and $m \neq 0$ in which case the $(-m)$-th moment is finite if

- *$\gamma_- = 1/2$ and $a_- > m\frac{\sigma_-^2}{2}$.*
- *$\gamma_- \in [0, 1/2)$ and $m < 1 - 2\gamma_- \leq 1$.*

Acknowledgements. The authors are grateful to P. Pigato for discussions and suggestions about this work. They also thank A. Lejay, and L. Lenôtre for references on existing bounds for first hitting times and C. Blanchet-Scalliet and D. Dorobantu for discussions on related topics.

References

[1] A. Alfonsi. On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo Methods Appl.*, 11(4):355–384, 2005.

[2] 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.

[3] R. F. Bass, K. Burdzy, and Z. Chen. Pathwise uniqueness for a degenerate stochastic differential equation. *Ann. Probab.*, 35(6):2385–2418, 2007.

[4] M. Ben Alaya and A. Kebaier. Parameter estimation for the square-root diffusions: Ergodic and nonergodic cases. *Stoch. Models*, 28(4):609–634, 2012.

[5] M. Ben Alaya and A. Kebaier. Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions. *Stochastic Anal. Appl.*, 31(4):552–573, 2013.

[6] A. N. Borodin and P. Salminen. *Handbook of Brownian motion: Facts and formulae.* Probab. Appl. Basel: Birkhäuser, 2nd ed. edition, 2002.

[7] R.-S. Cantrell and C. Cosner. Diffusion models for population dynamics incorporating individual behavior at boundaries: applications to refuge design. *Theoretical population biology*, 55(2):189–207, 1999.

[8] K. C. Chan, G. Karolyi, F. Longstaff, and A. Sanders. The volatility of short-term interest rates: an empirical comparison of alternative models of the term structure of interest rates. In *The new interest rate models. Recent developments in the theory and application of yield curve dynamics*, pages 87–100. London: Risk Books, 2000.

[9] I. Crimaldi and L. Pratelli. Convergence results for multivariate martingales. *Stochastic Processes Appl.*, 115(4):571–577, 2005.

[10] M. Decamps, M. Goovaerts, and W. Schoutens. Self exciting threshold interest rates models. *Int. J. Theor. Appl. Finance*, 9(7):1093–1122, 2006.

[11] H. J. Engelbert and W. Schmidt. Strong Markov continuous local martingales and solutions of one- dimensional stochastic differential equations. III. *Math. Nachr.*, 151:149–197, 1991.

[12] Y. Hu and Y. Xi. Parameter estimation for threshold Ornstein–Uhlenbeck processes from discrete observations. *Journal of Computational and Applied Mathematics*, 411:114264, 2022.

[13] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Math. Libr.* Elsevier (North-Holland), Amsterdam, 1981.

[14] J. Jacod and P. Protter. *Discretization of processes.*, volume 67 of *Stoch. Model. Appl. Probab.* Berlin: Springer, 2012.

[15] J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes.*, volume 288 of *Grundlehren Math. Wiss.* Berlin: Springer, 2nd ed. edition, 2003.

[16] O. Kallenberg. *Foundations of modern probability.* Probab. Appl. New York, NY: Springer, 2nd ed. edition, 2002.

[17] Y. A. Kutoyants. On identification of the threshold diffusion processes. *Ann. Inst. Statist. Math.*, 64(2):383–413, 2012.

[18] L. Le Cam and G. Lo Yang. *Asymptotics in statistics: some basic concepts.* Springer Ser. Stat. New York etc.: Springer-Verlag, 1990.

[19] J. F. Le Gall. One-dimensional stochastic differential equations involving the local times of the unknown process. Stochastic analysis and applications, Proc. int. Conf., Swansea 1983, Lect. Notes Math. 1095, 51-82 (1984)., 1984.

[20] A. Lejay and P. Pigato. Statistical estimation of the oscillating Brownian motion. *Bernoulli*, 24(4B):3568–3602, 2018.

[21] 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.

[22] A. Lejay and P. Pigato. Maximum likelihood drift estimation for a threshold diffusion. *Scand. J. Stat.*, 47(3):609–637, 2020.

[23] D. Lepingle. Sur le comportement asymptotique des martingales locales. Semin. Probab. XII, Univ. Strasbourg 1976/77, Lect. Notes Math. 649, 148-161 (1978)., 1978.

[24] S. Mazzonetto and P. Pigato. Drift estimation of the threshold Ornstein-Uhlenbeck process from continuous and discrete observations. *Stat. Sin.*, 34(1):313–336, 2024.

[25] S. Mazzonetto and P. Pigato. Estimation of parameters and local times in a discretely observed threshold diffusion model. *Preprint arXiv:2403.06858*, 2024.

[26] Y. Mishura, K. Ralchenko, and O. Dehtiar. Parameter estimation in CKLS model by continuous observations. *Stat. Probab. Lett.*, 184:10, 2022. Id/No 109391.

[27] Y. Mishura and A. Yurchenko-Tytarenko. Standard and fractional reflected Ornstein-Uhlenbeck processes as the limits of square roots of Cox-Ingersoll-Ross processes. *Stochastics*, 95(1):99–117, 2023.

[28] P. P. Mota and M. L. Esquível. On a continuous time stock price model with regime switching, delay, and threshold. *Quant. Finance*, 14(8):1479–1488, 2014.

[29] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren Math. Wiss.* Berlin: Springer, 3rd ed., 3rd. corrected printing edition, 2005.

[30] F. Su and K. Chan. Quasi-likelihood estimation of a threshold diffusion process. *J. Econom.*, 189(2):473–484, 2015.

[31] F. Su and K.-S. Chan. Testing for threshold diffusion. *Journal of Business & Economic Statistics*, 35(2):218–227, 2017.

[32] D. Taguchi and A. Tanaka. On the Euler–Maruyama scheme for degenerate stochastic differential equations with non-sticky condition. *Séminaire de Probabilités L*, pages 165–185, 2019.

[33] T.-H. Yu, H. Tsai, and H. Rachinger. Approximate maximum likelihood estimation of a threshold diffusion process. *Computational Statistics & Data Analysis*, 142:106823, 2020.