Statistica Sinica Preprint No: SS-2021-0275	
Title	Drift Estimation of the Threshold Ornstein-Uhlenbeck Process From Continuous and Discrete Observations
Manuscript ID	SS-2021-0275
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0275
Complete List of Authors	Sara Mazzonetto and Paolo Pigato
Corresponding Authors	Paolo Pigato
E-mails	paolo.pigato@uniroma2.it

Statistica Sinica

Drift estimation of the threshold Ornstein–Uhlenbeck process from continuous and discrete observations

Sara Mazzonetto and Paolo Pigato

Université de Lorraine and University of Rome Tor Vergata

Abstract: The *threshold Ornstein–Uhlenbeck* process is a continuous-time threshold autoregressive process. It follows the Ornstein–Uhlenbeck dynamics when above or below a fixed threshold, but its coefficients can be discontinuous at the threshold. We discuss (quasi)-maximum likelihood estimation of the drift parameters, assuming continuous and discrete time observations. In the ergodic case, we derive the consistency and the speed of convergence of these estimators in long time and high frequency. Based on these results, we develop a test for the presence of a threshold in the dynamics. Finally, we apply these statistical tools to short-term US interest rates modeling.

Key words and phrases: Threshold diffusion, maximum likelihood, regime-switching, self-exciting process, interest rates, threshold Vasicek Model, multi-threshold.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

1. Introduction

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

$$X_t = X_0 + \int_0^t \sigma(X_s)\,\mathrm{d}W_s + \int_0^t \left(b(X_s) - a(X_s)\,X_s\right)\,\mathrm{d}s, \quad t \geq 0, \tag{1.1}$$

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

$$\sigma(x) = \sigma_+ \mathbf{1}_{\{x \geq r\}} + \sigma_- \mathbf{1}_{\{x < r\}} > 0, \tag{1.2}$$

and a similarly piecewise affine drift coefficient

$$b(x) = b_+ \mathbf{1}_{\{x \geq r\}} + b_- \mathbf{1}_{\{x < r\}} \quad \text{and} \quad a(x) = a_+ \mathbf{1}_{\{x \geq r\}} + a_- \mathbf{1}_{\{x < r\}}. \tag{1.3}$$

The strong existence of a unique solution to (1.1) follows from the results of Le Gall (1985). Separately on (r, ∞) and $(-\infty, r)$, the process follows the Ornstein–Uhlenbeck (OU) dynamics, which, in the context of interest rate modeling, is referred to as the *Vasicek model*. Following this nomenclature, Decamps et al. (2006) refer to (1.1) as the *self–exciting threshold (SET) Vasicek model*, and Su and Chan (2015, 2017) refer to it as the *threshold diffusion* (TD) or first-order continuous–time threshold autoregressive (TAR) model (see also (Tong, 1990)).

The process is *ergodic* if the drift pushes the process up when the process reaches large, negative values, and down when it reaches large, positive

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

values. Note that if $a_- > 0$, the drift points toward b_-/a_- when X_s is below the threshold r, and if $a_+ > 0$, it points toward b_+/a_+ when X_s is above r. Here, we allow a null linear part, and if $a_- = 0$, we need $b_- > 0$ to push the process up when very negative, and if $a_+ = 0$, we need $b_+ < 0$ to push it down when very positive. Based on these considerations, we can easily check for ergodicity (see condition (2.1)).

In Section S3 of the Supplementary Material, we also consider a *multi-threshold* version of (1.1), where we allow for d discontinuity levels $r_1 < \cdots < r_d$. In this case, ergodicity is determined by the same conditions, checked on the values of the coefficients on the intervals $(-\infty, r_1)$ and $[r_d, +\infty)$.

In this paper, we discuss the asymptotic behavior of maximum likelihood estimators (MLE) and quasi-maximum likelihood estimators (QMLE) for the drift parameters (a_-, a_+, b_-, b_+), from both continuous and discrete time observations. Let N be the number of observations, T_N be the time horizon, and Δ_N be the largest interval between two consecutive observations. In the ergodic case, if $T_N \to \infty$ and $T_N \Delta_N \to 0$ as $N \to \infty$, we prove a central limit theorem (CLT) in which the estimators converge with speed $\sqrt{T_N}$ to the real parameters, that is, asymptotic normality (see Theorem 2). To the best of our knowledge, this is the first result of this kind for TDs

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

(SDEs with discontinuous drift and diffusion coefficients).

The discontinuity in the coefficients makes it difficult to pass from discrete-time to continuous-time observations. Indeed, a precise analysis of the error hinges on the behavior of certain discretizations of the local time of the diffusion at the threshold. We also prove, for a fixed time horizon, that the discrete (Q)MLE based on N equally spaced observations converges in high frequency to the continuous (Q)MLE, with speed $N^{1/4}$ (see Theorem 3). This slow convergence of the discrete (Q)MLE to the continuous (Q)MLE follows from the slow convergence, with speed $N^{1/4}$, of the discretization of the local time.

Based on these results, we provide a test to decide whether a threshold is present in the dynamics. Finally, we use these tools to analyze short-term US interest rates.

Literature review. Su and Chan (2015, 2017) study the asymptotic behavior of the continuous-time QMLE of a TD with drift as in (1.3) and piecewise regular diffusivity. In particular, they construct a hypothesis test to decide whether the drift is affine or piecewise affine.

(Lejay and Pigato, 2018) estimate the volatility parameters $\sigma_{\pm}$ in (1.1) from high-frequency data, and (Lejay and Pigato, 2020) examine the drift

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

estimation in the case $a_\pm = 0$. In the purely linear drift case $b_\pm = 0$, Kutoyants (2012) studies the problem of identifying the threshold parameter r, and Dieker and Gao (2013), among others, consider similar models with $r = 0$ (so that the drift function is continuous) in a multidimensional setting. The (related) problem of drift estimation in a skew OU process is considered by (Xing et al., 2020).

In this study the coefficients are discontinuous and the behavior at r is difficult to handle; for high-frequency observations, we do so using the discretization results of (Mazzonetto, 2019). The convergence in high frequency and long time for estimators of discretely observed diffusions have been discussed, for example, in (Kessler, 1997; Ben Alaya and Kebaier, 2013; Amorino and Gloter, 2020), but to the best of our knowledge, ours is the first such result in the case of discontinuous coefficients.

Yu et al. (2020) numerically study an approximate MLE (AMLE) from discrete-time observations simultaneously for threshold, drift, and diffusion coefficients of threshold diffusions, including the OU process and the Cox-Ingersoll-Ross (CIR) model. They compare their AMLE with the QMLE, showing numerical evidence of consistency. (Hu and Xi, 2022) consider a generalized moment estimator for a TD that is observed discretely, with a fixed time lag.

TAR models in discrete time were introduced by H. Tong in the early 1980s (Tong, 1983, 2011, 2015). Within this class, self-exciting TAR (SE-TAR) models rely on a spatial segmentation, with the dynamics changing according to the position of the process, below or above a threshold, and can be viewed as a discrete analogue of the TD; see (Chan, 1993; Rabemananjara and Zakoian, 1993; Yadav et al., 1994; Brockwell and Williams, 1997; Chen et al., 2011), and the references therein, for this class of econometric models and related inference problems.

Diffusion processes are widely used to model interest rate dynamics: see, for example, (Vasicek, 1977; Cox et al., 1985; Hull and White, 1990; Black and Karasinski, 1991). These models are designed to capture the fact that interest rates are typically mean reverting (Wu and Zhang, 1996), but do not capture nonlinear effects (e.g., multi-modality). Ait-Sahalia (1996) shows that mean-reversion for interest rates is strong outside a middle region, suggesting the existence of a target band. This is similar to what is observed in exchange rates (Krugman, 1991), and can be explained by policy adjustments in response to changes in such rates. There is evidence of a "normal" low-mean regime and an "exceptional" high-mean regime and, in general, bi-modality (or even multi-modality) in interest rate dynamics, which we can model using a TD (1.3). In general, nonlinearities and regime

changes in short-term interest rates are widely documented, and several discrete-time and continuous-time threshold models have been proposed; see (Gray, 1996; Pfann et al., 1996; Ang and Bekaert, 2002a,b; Kalimipalli and Susmel, 2004; Gospodinov, 2005; Ang et al., 2008; Archontakis and Lemke, 2008a,b). See also (Decamps et al., 2006) and the references therein for a thorough discussion of SET diffusions in interest rate modeling. TDs have recently been used in several aspects of financial modeling, such as option pricing (Lipton and Sepp, 2011; Gairat and Shcherbakov, 2016; Dong and Wong, 2017; Lipton, 2018; Pigato, 2019) and time series modeling (Ang and Timmermann, 2012; Lejay and Pigato, 2019). TD models for interest rates are considered in (Pai and Pedersen, 1999; Decamps et al., 2006; Su and Chan, 2015, 2017). In this study, we focus on (Q)MLE estimators of such models. In particular, we examine estimators from high-frequency observations and their convergence to continuous-time estimators, as well as their convergence in long time to the real values of the parameters.

In Section 2, we present our main results on the convergence of drift estimators for a threshold OU. In Section 3, we implement the estimators, discuss threshold estimation, and apply our work to US interest rate data. Proofs are collected in Section 4 and in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

2. (Quasi) maximum likelihood estimation

Equation (1.1), where W is a Brownian motion and X_0 is independent of W (e.g., X_0 is deterministic), with coefficients as in (1.2) and (1.3), admits a unique strong solution. Let X be such a process. We show in equation (4.3) that X is ergodic if

$$\begin{aligned}&[(a_+ > 0 \text{ and } b_+ \in \mathbb{R}) \text{ or } (a_+ = 0 \text{ and } b_+ < 0)]\\ &\text{and } [(a_- > 0 \text{ and } b_- \in \mathbb{R}) \text{ or } (a_- = 0 \text{ and } b_- > 0)].\end{aligned} \tag{2.1}$$

See Section S3 of the Supplementary Material for the analogous condition in the multi-threshold case.

2.1 MLE and QMLE from continuous-time observations

In this section, we observe a process on the time interval $[0, T]$, for $T \in (0, \infty)$. For $T \in (0, \infty)$, $m = 0, 1, 2$, and $\pm \in \{-, +\}$; we define

$$\mathfrak{M}_T^{\pm,m} := \int_0^T X_s^m \mathbf{1}_{\{\pm(X_s - r) \geq 0\}} \, \mathrm{d}X_s \quad \text{and} \quad \mathfrak{Q}_T^{\pm,m} := \int_0^T X_s^m \mathbf{1}_{\{\pm(X_s - r) \geq 0\}} \, \mathrm{d}s, \tag{2.2}$$

and take as the likelihood function the Girsanov weight

$$\begin{aligned}&G_T(a_+, b_+, a_-, b_-)\\ &= \exp\left(\int_0^T \frac{b(X_s) - a(X_s)X_s}{(\sigma(X_s))^2} \, \mathrm{d}X_s - \frac{1}{2}\int_0^T \frac{(b(X_s) - a(X_s)X_s)^2}{(\sigma(X_s))^2} \, \mathrm{d}s\right).\end{aligned} \tag{2.3}$$

We also consider the quasi-likelihood defined in (Su and Chan, 2015),

$$\Lambda_T(a_+, b_+, a_-, b_-) = \int_0^T b(X_s) - a(X_s)X_s \,\mathrm{d}X_s - \frac{1}{2}\int_0^T (b(X_s) - a(X_s)X_s)^2 \,\mathrm{d}s. \tag{2.4}$$

Theorem 1. *Let* $\pm \in \{+, -\}$.

i) For every $T \in (0, \infty)$*, provided that* $\mathfrak{Q}_T^{-,0} > 0$ *and* $\mathfrak{Q}_T^{+,0} > 0$*, the MLE and QMLE are given by*

$$\left(\alpha_T^{\pm}, \quad \beta_T^{\pm}\right) = \left(\frac{\mathfrak{M}_T^{\pm,0}\mathfrak{Q}_T^{\pm,1} - \mathfrak{Q}_T^{\pm,0}\mathfrak{M}_T^{\pm,1}}{\mathfrak{Q}_T^{\pm,0}\mathfrak{Q}_T^{\pm,2} - (\mathfrak{Q}_T^{\pm,1})^2}, \quad \frac{\mathfrak{M}_T^{\pm,0}\mathfrak{Q}_T^{\pm,2} - \mathfrak{Q}_T^{\pm,1}\mathfrak{M}_T^{\pm,1}}{\mathfrak{Q}_T^{\pm,0}\mathfrak{Q}_T^{\pm,2} - (\mathfrak{Q}_T^{\pm,1})^2}\right). \tag{2.5}$$

Assume now that the process is ergodic (i.e., (2.1) is satisfied).

ii) The following law of large numbers (LLN) holds: $(\alpha_T^{\pm} - a_{\pm}, \ \beta_T^{\pm} - b_{\pm}) \xrightarrow[T\to\infty]{\text{a.s.}} 0$*, that is, the estimator is consistent.*

iii) The following CLT jointly holds for the positive and negative sides:

$$\sqrt{T}\left(\alpha_T^{\pm} - a_{\pm}, \quad \beta_T^{\pm} - b_{\pm}\right) \xrightarrow[T\to\infty]{\text{stably}} N^{\pm} = \left(N^{\pm,\alpha}, \quad N^{\pm,\beta}\right),$$

where $\left(N^{+,\alpha}, \quad N^{+,\beta}\right)$ *and* $\left(N^{-,\alpha}, \quad N^{-,\beta}\right)$ *are two independent two-dimensional Gaussian random variables, independent of* X*, with covariance matrices* $\sigma_+^2\Gamma_+^{-1}$ *and* $\sigma_-^2\Gamma_-^{-1}$*, respectively, where*

$$\Gamma_{\pm} := \begin{pmatrix} \overline{\mathfrak{Q}}_{\infty}^{\pm,2} & -\overline{\mathfrak{Q}}_{\infty}^{\pm,1} \\ -\overline{\mathfrak{Q}}_{\infty}^{\pm,1} & \overline{\mathfrak{Q}}_{\infty}^{\pm,0} \end{pmatrix}, \tag{2.6}$$

and $\overline{\mathfrak{Q}}_{\infty}^{\pm,i}$*, for* $i \in \{0,1,2\}$*, are real constants such that* $\lim_{t\to\infty} \frac{\mathfrak{Q}_t^{\pm,i}}{t} \overset{\text{a.s.}}{=} \overline{\mathfrak{Q}}_{\infty}^{\pm,i}$ *(explicit expressions for such constants are given in Lemma 2; further details on the stable convergence are provided in Remark 1).*

iv) The local asymptotic normality *(LAN) property (see Le Cam and Yang (2000)) holds for the likelihood evaluated at the true parameter* $\theta := (a_+, b_+, a_-, b_-)$*, with rate of convergence* $\frac{1}{\sqrt{T}}$ *and asymptotic Fisher information*

$$\Gamma = \begin{pmatrix} \sigma_+^{-2}\Gamma_+ & 0_{\mathbb{R}^{2\times 2}} \\ 0_{\mathbb{R}^{2\times 2}} & \sigma_-^{-2}\Gamma_- \end{pmatrix}.$$

This means that there exists a random vector $A_T \in \mathbb{R}^4$ *such that for all small perturbations* $\Delta\theta := (\Delta a_+, \Delta b_+, \Delta a_-, \Delta b_-)$*, the quantity*

$$\log \frac{G_T(\theta + \frac{1}{\sqrt{T}}\Delta\theta)}{G_T(\theta)} - \left(\Delta\theta \cdot A_T - \frac{1}{2}\Delta\theta \cdot \Gamma\Delta\theta\right) \tag{2.7}$$

converges to zero in probability as $T \to \infty$*.*

Remark 1. The notion of *stable convergence* was introduced by Rényi (1963), and is discussed in detail by, among others, (Jacod and Shiryaev, 2003) or (Jacod and Protter, 2012). Here, we simply mention that this notion of convergence is stronger than convergence in law, but weaker than convergence in probability. We use the following crucial property: for ran-

dom variables Y_n, Z_n $(n \geq 1)$, Y and Z,

$$\text{if } Z_n \xrightarrow[n\to\infty]{\text{stably}} Z \text{ and } Y_n \xrightarrow[n\to\infty]{\mathbb{P}} Y, \text{ then } (Y_n, Z_n) \xrightarrow[n\to\infty]{\text{stably}} (Y, Z). \tag{2.8}$$

Remark 2. Considering continuous-time observations may seem in conflict with observed empirical time series, which always consist of "digital" data. However, it is common practice to compute the counterpart to continuous-time quantities from discrete data, for example approximating the integrals in (2.2) using the Euler–Maruyama scheme or Riemann sums; see (Su and Chan, 2015, 2017). This is usually better justified when the process is observed in high frequency, a closer setting to the continuous-time idealization. In the present study, continuous-time Theorem 1 can be applied using this strategy. Theorem 1 also serves as an intermediate step in the proof of Theorem 2, its discrete-time high frequency version.

2.2 Drift estimation from discrete observations

In this section we observe the process on a discrete time grid $0 = t_0 < t_1 < \ldots < t_{N-1} < t_N = T$, for $N \in \mathbb{N}$, $T \in (0, \infty)$, and set $\Delta_N = \max_{k=1,\ldots,N}\{t_k - t_{k-1}\}$. We define $X_i := X_{t_i}$ with $i = 0, \ldots, N$.

The discrete versions of (2.2) are defined as follows: for $m = 0, 1, 2$ and

$\pm \in \{-,+\}$, let

$$\begin{aligned}\mathfrak{M}_{T,N}^{\pm,m} &:= \sum_{k=0}^{N-1} X_k^m \mathbf{1}_{\{\pm(X_k-r)\geq 0\}}(X_{k+1}-X_k), \quad \text{and} \\ \mathfrak{Q}_{T,N}^{\pm,m} &:= \sum_{k=0}^{N-1} X_k^m \mathbf{1}_{\{\pm(X_k-r)\geq 0\}}(t_{k+1}-t_k).\end{aligned} \tag{2.9}$$

We refer to the *discretized likelihood* (corresponding to (2.3)) as

$$\begin{aligned}&G_{T,N}(a_+,b_+,a_-,b_-) \\ &= \exp\left(\sum_{i=0}^{N-1}\left(\frac{b(X_i)-a(X_i)X_i}{\sigma(X_i)^2}(X_{i+1}-X_i) - \frac{t_{i+1}-t_i}{2}\frac{(b(X_i)-a(X_i)X_i)^2}{\sigma(X_i)^2}\right)\right),\end{aligned}$$

and to the *discretized quasi-likelihood* (corresponding to (2.4)) as

$$\begin{aligned}&\Lambda_{T,N}(a_+,b_+,a_-,b_-) \\ &= \sum_{i=0}^{N-1}\left((b(X_i)-a(X_i)X_i)(X_{i+1}-X_i) - \frac{t_{i+1}-t_i}{2}(b(X_i)-a(X_i)X_i)^2\right).\end{aligned}$$

For $N \in \mathbb{N}$, $T \in (0,\infty)$, let

$$\left(\widehat{a}_{T,N}^{\pm},\ \widehat{b}_{T,N}^{\pm}\right) = \left(\frac{\mathfrak{M}_{T,N}^{\pm,0}\mathfrak{Q}_{T,N}^{\pm,1}-\mathfrak{Q}_{T,N}^{\pm,0}\mathfrak{M}_{T,N}^{\pm,1}}{\mathfrak{Q}_{T,N}^{\pm,0}\mathfrak{Q}_{T,N}^{\pm,2}-(\mathfrak{Q}_{T,N}^{\pm,1})^2},\ \frac{\mathfrak{M}_{T,N}^{\pm,0}\mathfrak{Q}_{T,N}^{\pm,2}-\mathfrak{Q}_{T,N}^{\pm,1}\mathfrak{M}_{T,N}^{\pm,1}}{\mathfrak{Q}_{T,N}^{\pm,0}\mathfrak{Q}_{T,N}^{\pm,2}-(\mathfrak{Q}_{T,N}^{\pm,1})^2}\right). \tag{2.10}$$

Theorem 2. *Let $(T_N)_{N\in\mathbb{N}}$ be a sequence in $(0,\infty)$. For all $N \in \mathbb{N}$, let $\widehat{a}_{T_N,N}^{\pm}, \widehat{b}_{T_N,N}^{\pm}$ be defined as in (2.10).*

i) For every $N \in \mathbb{N}$, provided that $\mathfrak{Q}_{T_N,N}^{-,0} > 0$ and $\mathfrak{Q}_{T_N,N}^{+,0} > 0$, the vector $(\widehat{a}_{T_N,N}^{+}, \widehat{b}_{T_N,N}^{+}, \widehat{a}_{T_N,N}^{-}, \widehat{b}_{T_N,N}^{-})$ maximizes both the likelihood $G_{T_N,N}(a_+,b_+,a_-b_-)$ and the quasi-likelihood $\Lambda_{T_N,N}(a_+,b_+,a_-b_-)$.

Assume now that the process is ergodic, that is, (2.1) *is satisfied, and that* X *is the stationary solution to* (1.1)*, that is,* X_0 *follows the stationary distribution (see* (4.2)*). Moreover, assume*

$$\lim_{N\to\infty} T_N = \infty \quad \textit{and} \quad \lim_{N\to\infty} \Delta_N = 0.$$

ii) The following LLN holds: $(\widehat{a}^{\pm}_{T_N,N},\ \widehat{b}^{\pm}_{T_N,N}) \xrightarrow[N\to\infty]{\mathbb{P}} (a_{\pm},\ b_{\pm})$ *(the estimator is consistent).*

iii) If $\lim_{N\to\infty} T_N\Delta_N = 0$*, the following CLT jointly holds for the positive and negative sides:*

$$\sqrt{T_N}\left(\widehat{a}^{\pm}_{T_N,N} - a_{\pm},\ \ \widehat{b}^{\pm}_{T_N,N} - b_{\pm}\right) \xrightarrow[N\to\infty]{\text{stably}} N^{\pm} = \left(N^{\pm,\alpha},\ \ N^{\pm,\beta}\right),$$

where $\left(N^{+,\alpha},\ \ N^{+,\beta}\right)$ *and* $\left(N^{-,\alpha},\ \ N^{-,\beta}\right)$ *are as in Theorem 1.*

iv) If $\lim_{N\to\infty} T_N\Delta_N = 0$*, the discretized likelihood satisfies* (2.7) *(with* $T = T_N$*).*

Remark 3. If the largest time lag $\Delta_N = O(T_N/N)$, the conditions in Theorem 2 become $\lim_{N\to\infty} T_N = \infty$ and $\lim_{N\to\infty} T_N/N = 0$ for consistency, and $\lim_{N\to\infty} T_N = \infty$ and $\lim_{N\to\infty} T_N^2/N = 0$ for asymptotic normality.

The next result states that, for fixed a time horizon, in high frequency, the estimator from discrete observations converges with an "anomalous"

speed toward the estimator from continuous observations. Let $Y\colon \Omega \times [0,\infty) \to \mathbb{R}$ be a semi-martingale, $r \in \mathbb{R}$, and $T \in [0,\infty)$. Then, we recall that

$$L_T^r(Y) = \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_0^T \mathbf{1}_{\{-\varepsilon \le Y_s - r \le \varepsilon\}} d\langle Y\rangle_s \tag{2.11}$$

is the symmetric local time of Y at r, up to time T.

Theorem 3. *Let $T \in (0,\infty)$ be fixed.*

i) For every $N \in \mathbb{N}$, provided that $\mathfrak{Q}^{-,0}_{T_N,N} > 0$ and $\mathfrak{Q}^{+,0}_{T_N,N} > 0$, the likelihood $G_{T,N}(a_+, b_+, a_-b_-)$ and the quasi-likelihood $\Lambda_{T,N}(a_+, b_+, a_-b_-)$ are both maximal at $(\widehat{a}^+_{T,N}, \widehat{b}^+_{T,N}, \widehat{a}^-_{T,N}, \widehat{b}^-_{T,N})$, given in (2.10).

ii) Assume the observations are equally spaced ($t_j = jT/N$, $j = 0,\dots,N$). It holds that $(\widehat{a}^+_{T,N}, \widehat{b}^+_{T,N}, \widehat{a}^-_{T,N}, \widehat{b}^-_{T,N}) \xrightarrow[N\to\infty]{\mathbb{P}} (\alpha^+_T, \beta^+_T, \alpha^-_T, \beta^-_T)$ and

$$\begin{aligned} &N^{1/4}\Big((\widehat{a}^+_{T,N}, \widehat{b}^+_{T,N}, \widehat{a}^-_{T,N}, \widehat{b}^-_{T,N}) - (\alpha^+_T, \beta^+_T, \alpha^-_T, \beta^-_T)\Big) \\ &\xrightarrow[N\to\infty]{\text{stably}} \sqrt{\frac{4\sqrt{T}}{3\sqrt{2\pi}} \frac{\sigma_-^2 + \sigma_+^2}{\sigma_- + \sigma_+}} \Bigg(\frac{\mathfrak{Q}^{+,1}_T - r\mathfrak{Q}^{+,0}_T}{\mathfrak{Q}^{+,0}_T\mathfrak{Q}^{+,2}_T - (\mathfrak{Q}^{+,1}_T)^2}, \frac{\mathfrak{Q}^{+,2}_T - r\mathfrak{Q}^{+,1}_T}{\mathfrak{Q}^{+,0}_T\mathfrak{Q}^{+,2}_T - (\mathfrak{Q}^{+,1}_T)^2}, \\ &\qquad - \frac{\mathfrak{Q}^{-,1}_T - r\mathfrak{Q}^{-,0}_T}{\mathfrak{Q}^{-,0}_T\mathfrak{Q}^{-,2}_T - (\mathfrak{Q}^{-,1}_T)^2}, - \frac{\mathfrak{Q}^{-,2}_T - r\mathfrak{Q}^{-,1}_T}{\mathfrak{Q}^{-,0}_T\mathfrak{Q}^{-,2}_T - (\mathfrak{Q}^{-,1}_T)^2} \Bigg) B_{L^r_T(X)}, \end{aligned} \tag{2.12}$$

with B a Brownian motion independent of X, and $L^r_T(X)$ the symmetric local time of X at r, up to time T (see (2.11)).

Remark 4. The right-hand side of (2.12) has the same law as

$$\sqrt{\frac{4\sqrt{T}}{3\sqrt{2\pi}}\frac{\sigma_-^2+\sigma_+^2}{\sigma_-+\sigma_+}}\begin{pmatrix}\begin{pmatrix}\mathfrak{Q}_T^{+,2} & -\mathfrak{Q}_T^{+,1}\\ -\mathfrak{Q}_T^{+,1} & \mathfrak{Q}_T^{+,0}\end{pmatrix}^{-1} & \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}\\ \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix} & \begin{pmatrix}\mathfrak{Q}_T^{-,2} & -\mathfrak{Q}_T^{-,1}\\ -\mathfrak{Q}_T^{-,1} & \mathfrak{Q}_T^{-,0}\end{pmatrix}^{-1}\end{pmatrix}\begin{pmatrix}-r\\ 1\\ r\\ -1\end{pmatrix}\sqrt{L_T^r(X)}B_1.$$

Remark 5. Theorem 1 and 2 are given as long-time limits, so that it is clear that as $T\to\infty$ the process visits both $[r,\infty)$ and $(-\infty,r)$. On the contrary, in Theorem 3 one needs to have observations of the process above and below the threshold before (the fixed) time T. If this is not true, say for example the process does not visit $[r,\infty)$ before time T, then the estimators (α_T^+,β_T^+) are not defined, even in the high-frequency limit $N\to\infty$.

Remark 6. One usually expects such discretizations to converge with speed $\sqrt{N}$. In this case, the lower speed of convergence is due to the discontinuity in the coefficients, and appears in connection with the local time. Indeed, the asymptotic behavior of the estimators is intrinsically related to that of the local time of the process at the threshold. More precisely, the difference $\mathfrak{M}_{T,N}^{\pm,m}-\mathfrak{M}_T^{\pm,m}$, for $m=0,1$, $\pm\in\{-,+\}$, can be rewritten involving the terms $L_{T,N}^r-L_T^r(X)$, where $L_{T,N}^r$ is the following approximation of the local time from discrete-time observations:

$$L_{T,N}^r:=2\sum_{i=0}^{N-1}\mathbf{1}_{\{(X_{iT/N}-r)(X_{(i+1)T/N}-r)<0\}}|X_{(i+1)T/N}-r|, \tag{2.13}$$

for $N \in \mathbb{N}$ (see equation (S2.6) in the Supplementary Material for a more precise statement).

Remark 7 (The skew OU process)**.** Consider the solution to the SDE involving the local time

$$Y_t = Y_0 + \int_0^t \bar{\sigma}(Y_s)\,\mathrm{d}W_s + \int_0^t \left(\bar{b}(Y_s) - \bar{a}(Y_s)\,Y_s\right)\,\mathrm{d}s + \bar{\beta} L_t^{\bar{r}}(Y), \quad t \geq 0, \tag{2.14}$$

with $\bar{\beta} \in (-1, 1)$, and some piecewise constant functions $\bar{\sigma}$, $\bar{a}$, and $\bar{b}$ possibly discontinuous at the threshold $\bar{r} \in \mathbb{R}$, as in (1.2) and (1.3).

Xing et al. (2020) assume $\bar{\beta}$ and $\bar{\sigma}$ are known, and estimate the drift parameters of Y, based on discrete observations, for constant $\bar{\sigma}, \bar{a}, \bar{b}$ coefficients and local time at zero. In this setting, Y is referred to as a "skew OU process" (see also (Feng, 2016)).

Consider now the more general case of $\bar{\sigma}, \bar{a}, \bar{b}$ in (1.2) and (1.3). If we assume that only $\bar{\beta}$ is known, the results in Section 2 for the drift estimation of X all hold similarly for the drift estimation of Y. This follows from the fact that a simple transformation allows us to reduce the skew OU to a threshold OU with threshold at zero, getting rid of the local time in the dynamics.

Remark 8 (The threshold CIR process)**.** Su and Chan (2015, 2017) and Yu et al. (2020) consider diffusion processes with drift as in (1.3), with more

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

flexibility on the diffusion coefficient $\sigma(\cdot)$, so that their analysis also apply to the process in (1.1), with

$$\sigma(x) = \sigma_+ \sqrt{x}\, \mathbf{1}_{\{x \geq r\}} + \sigma_- \sqrt{x}\, \mathbf{1}_{\{x < r\}}. \tag{2.15}$$

We refer to this process as the threshold CIR process. In the aformentioned works, the proposed estimators are always QMLE, maximizing (2.4), which does not depend on the diffusivity $\sigma(\cdot)$. Here, with our (more restrictive) piecewise-constant choice for the diffusivity, we show that the considered estimator is a genuine MLE. In our setting, we expect a result analogous to Theorem 1 to apply to the QMLE for the threshold CIR as well, but with a less explicit limit Gaussian law in the CLT; see also (Su and Chan, 2015). On the other hand, the proof of the discrete-time Theorem 2 uses bounds on the hitting times for the OU process. The corresponding result for the threshold CIR process is not a trivial extension.

3. Threshold estimation, testing and interest rates

We simulate the threshold OU process using the Euler scheme (Bokil et al., 2020) (an alternative approach is to discretize space instead of time; see (Ding et al., 2020)), and use the estimator based on discrete observations. The implementation is performed using `R`, and the parameters are as in Table 1. Figure 1 shows an example of the CLT in Theorem 2. In Table

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

r	b_-	b_+	a_-	a_+	σ_-	σ_+
0.01	-0.002	0.003	0.1	0.11	0.011	0.01

Table 1: Simulations parameters.

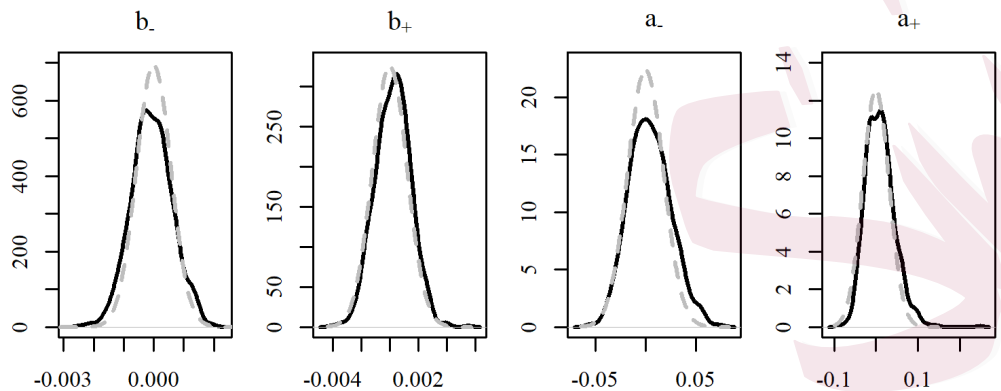

Figure 1: CLT in Theorem 2.(iii), with parameters as in Table 1. We plot the density of the theoretical distribution of the estimation error (dashed line) and compare it with the distribution of the error on $n = 10^3$ trajectories, with $T = 10^3$ and $N = 10^6$ observations on each trajectory.

2, we show the mean, standard deviation, and mean squared error of the estimators on simulated trajectories. Figure 2 shows an example of the convergence in Theorem 3, with (2.12) rewritten as

$$N^{1/4}\Big(\frac{\widehat{a}^{\pm}_{T,N} - \alpha^{\pm}_T}{\mathfrak{Q}^{\pm,1}_T - r\mathfrak{Q}^{\pm,0}_T}, \frac{\widehat{b}^{\pm}_{T,N} - \beta^{\pm}_T}{\mathfrak{Q}^{\pm,2}_T - r\mathfrak{Q}^{\pm,1}_T}\Big)\sqrt{\frac{3\sqrt{2\pi}}{4\sqrt{T}L^r_T(X)}\frac{\sigma_- + \sigma_+}{\sigma_-^2 + \sigma_+^2}(\mathfrak{Q}^{\pm,0}_T\mathfrak{Q}^{\pm,2}_T - (\mathfrak{Q}^{\pm,1}_T)^2)}$$
$$\xrightarrow[N\to\infty]{\text{stably}} \pm B_1, \text{ for } \pm \in \{-,+\}. \tag{3.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

parameter	b_-	b_+	a_-	a_+
mean	-0.00204	0.00318	0.105	0.119
MSE	5.10×10^{-7}	1.62×10^{-6}	0.000526	0.00130
simulated sd	0.000713	0.00126	0.0223	0.0349
predicted sd	0.000575	0.00122	0.0178	0.0317

Table 2: Mean, mean squared error (MSE), and standard deviation (simulated sd) of the (Q)MLE estimators in (2.10), with parameters as in Table 1, on $n = 10^3$ trajectories, with $T = 10^3$ and $N = 10^6$ observations on each trajectory. The "predicted sd" is the sd predicted by the Gaussian CLT in Theorem 2.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

To estimate the local time $L_T^r(X)$ and the occupation times $\mathfrak{O}_T^{\pm,i}$, we use the discrete-time approximations in (2.13) and (2.9), respectively. To simulate

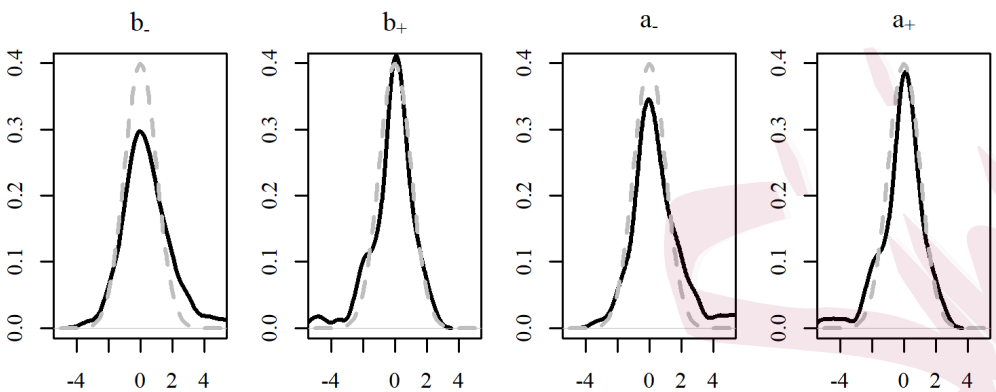

Figure 2: Convergence in Theorem 3, with parameters as in Table 1. We compare the empirical distribution on the left-hand side of (3.1) (on $n = 100$ simulated trajectories on the time interval $[0, 1]$, with initial condition $X_0 = r = 0$ and $N = 500$ discrete observations for each trajectory) with a standard Gaussian density (dashed line).

a stationary version of process (1.1), we can simulate X_0 using the explicit stationary density (4.2), or run the process until a large time T, and use X_T as the initial condition. In Figure 3, we compare the empirical distribution obtained in this way with the theoretical stationary density. This is an example of a bi-modal stationary distribution (density) with two peaks, corresponding to the two different mean reversion levels.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

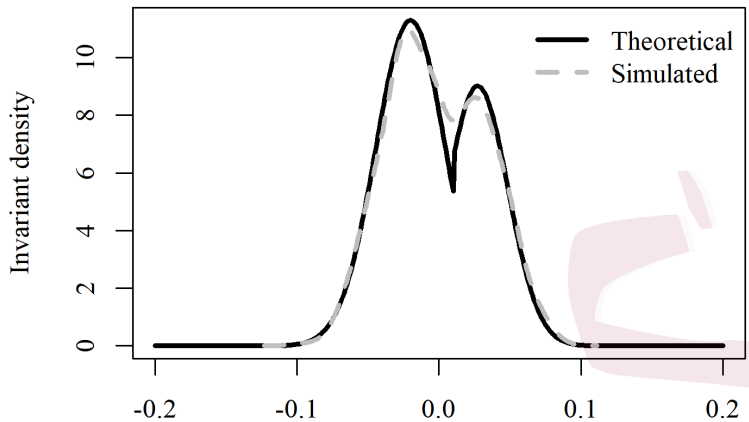

Figure 3: Theoretical invariant density in (4.2) versus the empirical distribution of X_T, with $T = 10^3$ and $N = 10^6$ discretization steps in the Euler scheme, on $n = 10^3$ trajectories. The parameters are as in Table 1.

3.1 On threshold estimation

The estimation results in Section 2 suppose we have previous knowledge of the threshold. In practice, this assumption is not realistic, and the threshold r usually has to be estimated. In (Su and Chan, 2015), the threshold QMLE from continuous observations is shown to be T-consistent. We implement the analogous threshold MLE, anddirectly consider discrete observations starting from the convergence results in Theorem 2.

Given N discrete observations of one trajectory up to time T_N, we proceed as follows. First, for a given threshold r, we compute (Q)MLE $(\widehat{a}^{\pm}, \widehat{b}^{\pm})_{T_N,N}$, and denote this estimator as $(\widehat{a}^{\pm}, \widehat{b}^{\pm})^{r}_{T_N,N}$. For each fixed r, we then compute the quasi-likelihood function $\Lambda_{T_N,N}$. We can also compute the likelihood function $G_{T_N,N}$, after estimating $\sigma_{\pm}$ using the quadratic variation estimators in (Lejay and Pigato, 2018). We take c as the δ-percentile and d as the $(1-\delta)$-percentile of the observed data (in the implementation we always take $\delta = 0.15$ and vary r on a discrete grid). Maximizing the (quasi-) likelihood function over $r \in [c, d]$, we obtain the (Q)MLE of the threshold, $\widehat{r}$. The estimator of all the drift parameters is then $(\widehat{r}, (\widehat{a}^{\pm}, \widehat{b}^{\pm})^{\widehat{r}}_{T_N,N})$.

We display a sample trajectory in Figure 4, together with the threshold estimated on that trajectory and the mean reversion levels. The estimated parameters are given in Table 3 (compare with the simulation parameters in Table 1). Note that the MLE and QMLE give the same parameter estimates once the threshold is fixed (Theorem 2.(i)). However, when also maximizing over the choice of the threshold, the MLE can account for a possible change in the volatility, which may yield a different choice of threshold. The model with different volatilities (SET Vasicek) is used by Decamps et al. (2006). Su and Chan (2015, 2017) use the QMLE, thus their drift estimator does not account for possible changes in volatility.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

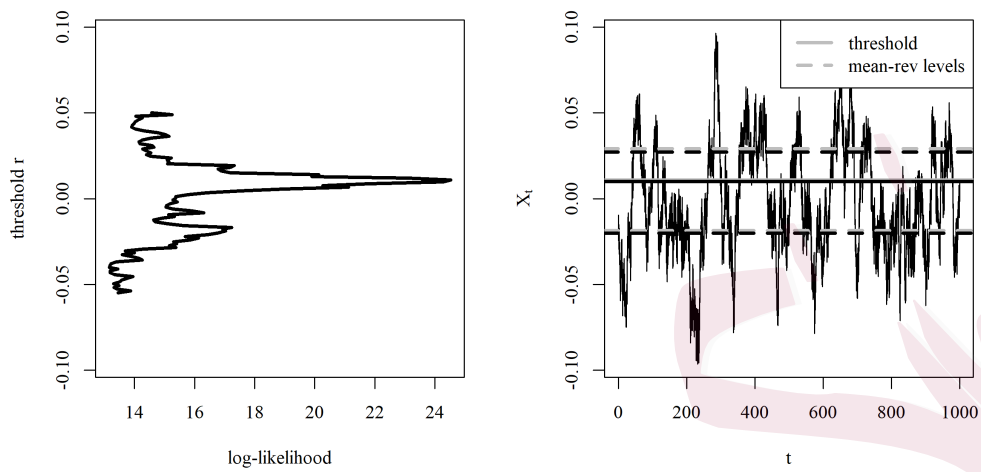

Figure 4: A sample trajectory with parameters as in Table 1, $T = 10^3$ and $N = 10^6$ time steps, and the results of estimating both the threshold and the parameters, using the MLE. On the left, we show the log-likelihood (on the x-axis) as a function of the threshold (on the y-axis), to visualize threshold estimation procedure described in Section 3.1. On the right, we show the estimated versus actual threshold levels and the mean reversion levels b_-/a_- and b_+/a_+.

3.2 Testing for threshold

Here, we test for the presence of a threshold in the diffusion dynamics. Su and Chan (2017) propose a test for the presence of a threshold based on the quasi-likelihood ratio. Here, we derive a test from the CLT in Theo-

parameter	r	b_-	b_+	a_-	a_+	b_-/a_-	b_+/a_+
(Q)MLE	0.0109	−0.00222	0.00403	0.119	0.138	−0.0186	0.0292
sd		0.000649	0.00129	0.0218	0.0342		

Table 3: Estimated drift parameters corresponding to Figure 4. Note that in this case, the threshold maximizing MLE and QMLE is the same (on the discrete grid we consider), but this is not necessarily the case. With the same threshold, the estimates for $a_\pm, b_\pm$ are the same, as follows from Theorem 2.(i). We also show the standard deviation of such estimators using to the Gaussian CLT in Theorem 2.

rem 2.(iii). Therefore, we suppose that its assumptions are satisfied. Moreover, we assume that the threshold parameter r is given. In practice, a natural choice for r is the (Q)MLE $\hat{r}$. With a fixed threshold, we can estimate the drift parameters, obtaining $(\widehat{a}^{\pm}_{T_N,N}, \widehat{b}^{\pm}_{T_N,N})$. From Theorem 2.(iii), if T_N^2/N goes to zero as $N \to \infty$,

$$\sqrt{T_N}\left((\widehat{a}^{+}_{T_N,N} - \widehat{a}^{-}_{T_N,N}) - (a_+ - a_-), (\widehat{b}^{+}_{T_N,N} - \widehat{b}^{-}_{T_N,N}) - (b_+ - b_-)\right) \xrightarrow[N\to\infty]{\text{stably}} N^+ - N^-, \tag{3.2}$$

which is a centered Gaussian vector with covariance matrix given by $\Sigma := \sigma_+^2\Gamma_+^{-1} + \sigma_-^2\Gamma_-^{-1}$, invertible from the Cauchy–Schwarz inequality. The inverse matrix Σ^{-1} can be expressed as a function of $\sigma_\pm$ and $\mathfrak{Q}_\infty^{\pm,i}$. Note that $\sigma_\pm$

can be estimated from one observed trajectory using quadratic variation, as in (Lejay and Pigato, 2018), and $\mathfrak{Q}_{\infty}^{\pm,i}$ can be estimated by computing $\frac{1}{T_N}\mathfrak{Q}_{T_N,N}^{\pm,i}$ as Riemann sums on the observed trajectory, from (2.6). We denote $\widehat{\Sigma}^{-1}$ as the estimate of Σ^{-1} obtained from these estimations.

To test for the presence of a threshold in the drift, we consider the hypothesis

$$\begin{cases} H_0\colon \text{ Null hypothesis} & (a_+, b_+) = (a_-, b_-) \\ H_1\colon \text{ Alternative hypothesis} & (a_+, b_+) \neq (a_-, b_-). \end{cases}$$

Under the null hypothesis, the statistic

$$T_N \left((\widehat{a}^+_{T_N,N} - \widehat{a}^-_{T_N,N}), (\widehat{b}^+_{T_N,N} - \widehat{b}^-_{T_N,N}) \right) \widehat{\Sigma}^{-1} \begin{pmatrix} (\widehat{a}^+_{T_N,N} - \widehat{a}^-_{T_N,N}) \\ (\widehat{b}^+_{T_N,N} - \widehat{b}^-_{T_N,N}) \end{pmatrix}$$

converges to a χ^2 distribution with two degrees of freedom. We reject H_0 if the statistic is larger than q_α, where q_α is the quantile of a χ^2 distribution with two degrees of freedom, such that $\mathbb{P}(\chi^2_2 \geq q_\alpha) = \alpha$.

To conclude, note that (3.2) allows us to test separately for the presence of a threshold in $a(\cdot)$ or in $b(\cdot)$, that is, testing for the presence of a discontinuity in the piecewise linear or in the piecewise constant, respectively, part of the drift.

3.3 Interest rate analysis

We consider the three-month US Treasury Bill rate, and examine the time series of daily closing rates between January 4, 1960 and April 29, 2020 (source: Yahoo Finance). We perform quasi-maximum and maximum likelihood estimation using (2.10), adopting the convention that the "daily" time interval is $\mathrm{d}t = 0.046$ months, with one month as the time unit. The number of observations is $N = 15057$, and $T \approx 60$ years. We choose $\delta = 0.15$ as the percentile for the search of the threshold, and we report both our MLE and QMLE parameters.

Figure 5 (bottom) shows that for the QMLE our result is consistent with that of (Su and Chan, 2015), and the estimation identifies two regimes. The first exhibits low rates, with negligible drift, so that the process is almost a martingale. In the high regime, the drift ensures a stronger reversion to lower rates when the rates are very high. In Figure 5 (top), we use MLE (with $\sigma_\pm$ estimated using quadratic variation), and the estimation identifies a low regime corresponding to the period of extremely low rates, with minimal fluctuations, following the 2007-2008. Almost all of the remaining part of the time series is in the high regime. The estimated volatilities are $\sigma_- = 0.186$ in the low regime, and $\sigma_+ = 0.453$ in the high regime. To compute the standard deviations of these estimates, we apply (Lejay and

Pigato, 2018, Corollary 3.8.), in the form of (Lejay and Pigato, 2019, Proposition 3.1), with the same justification the latter. We obtain a the standard deviation of 0.00472 for σ_-^2 and 0.0120 for σ_+^2. Using this MLE for the drift, the mean-reverting effect looks non-negligible, both above and below the threshold. Note that the parameter estimates obtained using the MLE and QMLE are substantially different. This is because of the different choice of threshold, which, in the QMLE, does not depend on the behavior of the volatility, but in the MLE, is influenced by the volatility as well. Thus, when using the MLE, one of the two regimes isolated by the threshold consists only of the period of extremely low and stable rates that followed the 2008 financial crisis.

parameter	r	b_-	b_+	a_-	a_+	b_-/a_-	b_+/a_+
MLE	0.919	0.0469	0.0492	0.284	0.0106	0.165	4.63
MLE sd		0.0223	0.0402	0.0757	0.00672		
QMLE	6.73	0.00131	0.417	0.00115	0.0481	1.14	8.67
QMLE sd		0.0341	0.144	0.00877	0.0153		

Table 4: Estimated drift parameters corresponding to Figure 5, and corresponding standard deviation for $b_\pm$ and $a_\pm$ from Theorem 2

In this analysis, following Su and Chan (2015, 2017), we estimated our

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

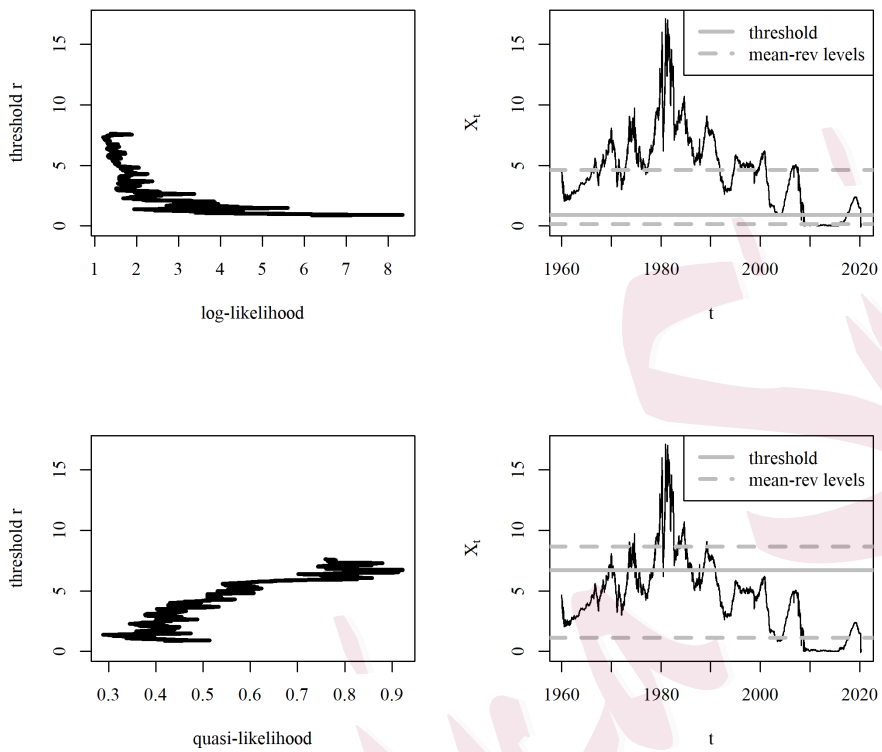

Figure 5: Three-month US Treasury Bill rate: time series of daily closing rates between January 4, 1960 and April 29, 2020. In the top figure, we use the MLE, and in the bottom figure, we use the QMLE. On the left-hand side, we show the log-likelihood and quasi-likelihood as functions of the threshold. On the right-hand side, we show estimated threshold levels (solid grey line), and the mean reversion levels b_-/a_- and b_+/a_+ (dashed grey line).

model parameters for the whole period 1960–2020. From an econometric perspective, it is natural to wonder whether it is reasonable to assume a stationary process on such a long time interval. To address this issue, we consider five two-years time windows, between 2010 and 2020, with daily observations as before. With this choice, $T^2 \approx N$, and therefore we expect from Theorem 2 that the discretization error should be negligible, assuming that $T \approx 2$ years is sufficiently large for the theorem to apply.

At a 1% significance level, we do not reject the H_0 hypothesis (absence of a threshold in the parameters) only in the subperiod January 2018-December 2019. In all other time windows (2010–2011, 2012–2013, 2014–2015, 2016–2017) we conclude that a threshold is present. Figure 6 shows three examples of estimations in such windows. In order to check whether this test is reliable for time series with such sample sizes, we tried the same procedure (selection of threshold and successive test with a 1% significance level) on simulated time series, with parameters and sample sizes of the same order as the estimated ones. We found that when no threshold is present (constant parameters), H_0 is rejected 14% of the time, and when the threshold is present (non-constant parameters) H_0 is rejected 96% of the time, which seems to confirm the validity of the procedure, even in these smaller time windows.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0275

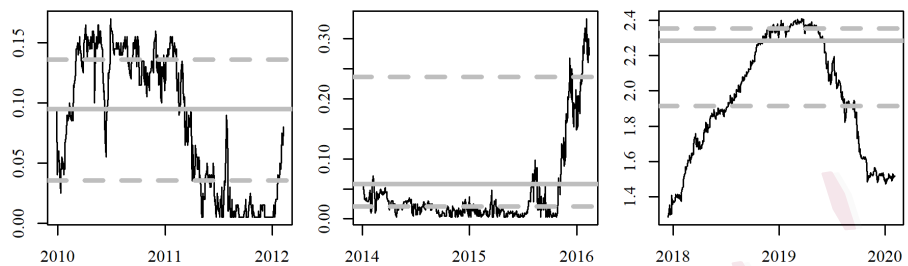

Figure 6: Three months US Treasury Bill rate: time series of daily closing rates for the periods January 2010 to December 2012, January 2014 to December 2016, January 2018 to December 2020. The estimated threshold level and mean reversion levels are given by the QMLE. Each time window consists of 24 months, with 22 observations per month. Our test concludes that a threshold in the dynamics is present in every time window, except for the period January 2018 to December 2020.

4. Proofs and technical results

4.1 The regimes of the process

In this section, we establish the values of the coefficients $(a_\pm, b_\pm)$ for which the process X is (positively or null) recurrent or transient. Because X is a one-dimensional diffusion, it is characterized by two quantities: *scale function* and *speed measure*, denoted by S and m, respectively.

Here, X is *recurrent* if and only if $\lim_{x\to+\infty} S(x) = +\infty$, and $\lim_{x\to-\infty} S(x) =$

$-\infty$, otherwise, it is transient. Moreover, a recurrent process is *positive recurrent* if the speed measure is a finite measure, otherwise it is null recurrent.

The scale function is continuous, unique up to multiplicative and additive constants, and its derivative satisfies (up to a multiplicative constant) $S'(x) = \exp\Big(-\int_r^x \frac{2(b(y)-a(y)y)}{(\sigma(y))^2}\,\mathrm{d}y\Big)$. We express the scale density explicitly as

$$s_\pm(x) := S'(x)\mathbf{1}_{\{\pm(x-r)\geq 0\}} = \exp\Big(-\frac{(x-r)}{\sigma_\pm^2}(2b_\pm - a_\pm(x+r))\Big). \tag{4.1}$$

It follows that X is recurrent if and only if $[(a_+ > 0 \text{ and } b_+ \in \mathbb{R}) \text{ or } (a_+ = 0 \text{ and } b_+ \leq 0)]$ and $[(a_- > 0 \text{ and } b_- \in \mathbb{R}) \text{ or } (a_- = 0 \text{ and } b_- \geq 0)]$. The complementary leads to transience.

The density of the *speed measure* with respect to the Lebesgue measure is given by $m(x) := \frac{2}{(\sigma(x))^2 S'(x)}$. It is *discontinuous* if and only if $\sigma^2(\cdot)$ is discontinuous. If X is recurrent and the speed measure is a finite measure, then X is positive recurrent and admits a stationary distribution, denoted by μ, which is equal to the renormalized speed measure

$$\mu(\mathrm{d}x) = \frac{m(x)}{\int_{-\infty}^{\infty} m(y)\,\mathrm{d}y}\,\mathrm{d}x. \tag{4.2}$$

In this case, we say that the process is *ergodic*. Lemma 1 states that the

speed measure is a finite measure when the following holds:

$$\begin{aligned}&[(a_+ > 0 \text{ and } b_+ \in \mathbb{R}) \text{ or } (a_+ = 0 \text{ and } b_+ < 0)]\\ &\text{and } [(a_- > 0 \text{ and } b_- \in \mathbb{R}) \text{ or } (a_- = 0 \text{ and } b_- > 0)].\end{aligned} \tag{4.3}$$

Under these conditions, the process is *ergodic*.

Lemma 1. *The speed measure is finite if and only if condition* (4.3) *holds. More precisely, let* $\pm \in \{-,+\}$ *and let*

$$\mathfrak{m}_\pm := \frac{\sqrt{\pi}}{\sigma_\pm\sqrt{a_\pm}} \exp\left(\frac{a_\pm}{\sigma_\pm^2}\left(\frac{b_\pm}{a_\pm} - r\right)^2\right) \operatorname{erfc}\left(\mp\frac{\sqrt{a_\pm}}{\sigma_\pm}\left(\frac{b_\pm}{a_\pm} - r\right)\right). \tag{4.4}$$

Then,

$$\int_{-\infty}^{\infty} \mathbf{1}_{\{\pm(y-r)\geq 0\}} m(y)\,\mathrm{d}y = \begin{cases} +\infty & \text{if } a_\pm = 0 \text{ and } b_\pm = 0,\\ \frac{1}{|b_\pm|} & \text{if } a_\pm = 0 \text{ and } \mp b_\pm > 0,\\ \mathfrak{m}_\pm & \text{if } a_\pm > 0 \text{ and } b_\pm \in \mathbb{R}.\end{cases}$$

In the following lemma, we use the ergodic property of the process to compute explicit expressions for the limits of suitably rescaled occupation times.

Lemma 2. *Assume the process is ergodic. Let* $\mathfrak{m}_-$ *and* $\mathfrak{m}_+$ *be given by* (4.4), $\mathfrak{b}_\pm = 1/|b_\pm|$, $\pm \in \{-,+\}$, *and* μ *be the stationary distribution of* X. *For all* $i \in \{0,1,2\}$, *let* $\overline{\mathfrak{Q}}_\infty^{\pm,i}$ *be the constant such that* $\overline{\mathfrak{Q}}_\infty^{\pm,i} \overset{\text{a.s.}}{=} \lim_{t\to\infty} \frac{\mathfrak{Q}_t^{i,\pm}}{t} \in \mathbb{R}$. *We have the following explicit formulae:*

- *if* $a_+ > 0$, $a_- > 0$, *and* $b_-, b_+ \in \mathbb{R}$, *then*

$$\overline{\mathfrak{Q}}_\infty^{\pm,0} = \frac{\mathfrak{m}_\pm}{\mathfrak{m}_+ + \mathfrak{m}_-}, \quad \overline{\mathfrak{Q}}_\infty^{\pm,1} = \frac{1}{\mathfrak{m}_+ + \mathfrak{m}_-}\left(\frac{b_\pm}{a_\pm}\mathfrak{m}_\pm \pm \frac{1}{a_\pm}\right), \quad \textit{and}$$
$$\overline{\mathfrak{Q}}_\infty^{\pm,2} = \frac{1}{\mathfrak{m}_+ + \mathfrak{m}_-}\left(\left(\frac{b_\pm^2}{a_\pm^2} + \frac{\sigma_\pm^2}{2a_\pm}\right)\mathfrak{m}_\pm \pm \left(\frac{b_\pm}{a_\pm} + r\right)\frac{1}{a_\pm}\right);$$

- *if* $a_+ = 0$, $a_- = 0$, *and* $b_+ < 0$, $b_- > 0$, *then*

$$\overline{\mathfrak{Q}}_\infty^{\pm,0} = \frac{\mathfrak{b}_\pm}{\mathfrak{b}_+ + \mathfrak{b}_-}, \quad \overline{\mathfrak{Q}}_\infty^{\pm,1} = \frac{\mathfrak{b}_\pm}{\mathfrak{b}_+ + \mathfrak{b}_-}\left(r \pm \frac{\sigma_\pm^2}{2}\mathfrak{b}_\pm\right), \quad \textit{and}$$
$$\overline{\mathfrak{Q}}_\infty^{\pm,2} = \frac{\mathfrak{b}_\pm}{\mathfrak{b}_+ + \mathfrak{b}_-}\left(r^2 \pm r\sigma_\pm^2\mathfrak{b}_\pm + \frac{\sigma_\pm^4}{2}(\mathfrak{b}_\pm)^2\right);$$

- *if* $a_+ > 0$, $b_+ \in \mathbb{R}$, $a_- = 0$, *and* $b_- > 0$, *then*

$$\overline{\mathfrak{Q}}_\infty^{+,0} = \frac{\mathfrak{m}_+}{\mathfrak{m}_+ + \mathfrak{b}_-}, \quad \overline{\mathfrak{Q}}_\infty^{-,0} = \frac{\mathfrak{b}_-}{\mathfrak{m}_+ + \mathfrak{b}_-}$$
$$\overline{\mathfrak{Q}}_\infty^{+,1} = \frac{1}{\mathfrak{m}_+ + \mathfrak{b}_-}\left(\frac{b_+}{a_+}\mathfrak{m}_+ + \frac{1}{a_+}\right), \quad \overline{\mathfrak{Q}}_\infty^{-,1} = \frac{\mathfrak{b}_-}{\mathfrak{m}_+ + \mathfrak{b}_-}\left(r - \frac{\sigma_-^2}{2}\mathfrak{b}_-\right),$$
$$\overline{\mathfrak{Q}}_\infty^{+,2} = \frac{1}{\mathfrak{m}_+ + \mathfrak{b}_-}\left(\left(\frac{b_+^2}{a_+^2} + \frac{\sigma_+^2}{2a_+}\right)\mathfrak{m}_+ + \left(\frac{b_+}{a_+} + r\right)\frac{1}{a_+}\right), \quad \textit{and}$$
$$\overline{\mathfrak{Q}}_\infty^{-,2} = \frac{\mathfrak{b}_-}{\mathfrak{m}_+ + \mathfrak{b}_-}\left(r^2 - r\sigma_-^2\mathfrak{b}_- + \frac{\sigma_-^4}{2}(\mathfrak{b}_-)^2\right);$$

- *if* $a_+ = 0$, $b_+ < 0$, $a_- > 0$, *and* $b_- \in \mathbb{R}$, *then*

$$\overline{\mathfrak{Q}}_\infty^{+,0} = \frac{\mathfrak{b}_+}{\mathfrak{b}_+ + \mathfrak{m}_-}, \quad \overline{\mathfrak{Q}}_\infty^{-,0} = \frac{\mathfrak{m}_-}{\mathfrak{b}_+ + \mathfrak{m}_-}$$
$$\overline{\mathfrak{Q}}_\infty^{+,1} = \frac{\mathfrak{b}_+}{\mathfrak{b}_+ + \mathfrak{m}_-}\left(r + \frac{\sigma_+^2}{2}\mathfrak{b}_+\right), \quad \overline{\mathfrak{Q}}_\infty^{-,1} = \frac{1}{\mathfrak{b}_+ + \mathfrak{m}_-}\left(\frac{b_-}{a_-}\mathfrak{m}_- - \frac{1}{a_-}\right),$$
$$\overline{\mathfrak{Q}}_\infty^{+,2} = \frac{\mathfrak{b}_+}{\mathfrak{b}_+ + \mathfrak{m}_-}\left(r^2 + r\sigma_+^2\mathfrak{b}_+ + \frac{\sigma_+^4}{2}(\mathfrak{b}_+)^2\right), \quad \textit{and}$$
$$\overline{\mathfrak{Q}}_\infty^{-,2} = \frac{1}{\mathfrak{b}_+ + \mathfrak{m}_-}\left(\left(\frac{b_-^2}{a_-^2} + \frac{\sigma_-^2}{2a_-}\right)\mathfrak{m}_- - \left(\frac{b_-}{a_-} + r\right)\frac{1}{a_-}\right).$$

4.2 Proof of Theorem 1

Proof of Item (i) *of Theorem 1.* Let $\theta := (a_+, b_+, a_-, b_-)$. It holds that

$$\Lambda_T(\theta) = \sum_{\pm\in\{-,+\}} \left(b_\pm \mathfrak{M}_T^{\pm,0} - a_\pm \mathfrak{M}_T^{\pm,1} - \frac{1}{2}\left(b_\pm^2 \mathfrak{Q}_T^{\pm,0} + a_\pm^2 \mathfrak{Q}_T^{\pm,2} - 2a_\pm b_\pm \mathfrak{Q}_T^{\pm,1} \right)\right). \tag{4.5}$$

To find the maximum, we compute the derivatives with respect to $a_\pm$ and $b_\pm$, and observe that the gradient has a unique singular point given by (2.5), and that the Hessian is negative definite.

Moreover, the fact that $\partial_{a_\pm}\Lambda_T = \sigma_\pm^2 \partial_{a_\pm} \log G_T$ and $\partial_{b_\pm}\Lambda_T = \sigma_\pm^2 \partial_{b_\pm} \log G_T$ shows that the MLE for the drift parameters is the same as the QMLE, that is, (2.5). □

In order to study the asymptotic behavior of the estimator, we introduce a different expression for the estimators in (2.5), based on the following notation. Given $T \in (0,\infty)$ and $i \in \{0,1\}$, let

$$M_T^{\pm,i} = \int_0^T (X_s)^i \mathbf{1}_{\{\pm(X_s - r)\geq 0\}} \, \mathrm{d}W_s.$$

Observe that (1.1) yields, for $i \in \{0,1\}$,

$$\mathfrak{M}_T^{\pm,i} = \sigma_\pm M_T^{\pm,i} + b_\pm \mathfrak{Q}_T^{\pm,i} - a_\pm \mathfrak{Q}_T^{\pm,i+1}. \tag{4.6}$$

Lemma 3. *Let $T \in (0, \infty)$. The MLE and QMLE can be expressed as*

$$\begin{cases} \alpha_T^{\pm} = a_{\pm} + \sigma_{\pm} \frac{M_T^{\pm,0}\mathfrak{Q}_T^{\pm,1} - \mathfrak{Q}_T^{\pm,0} M_T^{\pm,1}}{\mathfrak{Q}_T^{\pm,0}\mathfrak{Q}_T^{\pm,2} - (\mathfrak{Q}_T^{\pm,1})^2}, \\ \beta_T^{\pm} = b_{\pm} + \sigma_{\pm} \frac{M_T^{\pm,0}\mathfrak{Q}_T^{\pm,2} - \mathfrak{Q}_T^{\pm,1} M_T^{\pm,1}}{\mathfrak{Q}_T^{\pm,0}\mathfrak{Q}_T^{\pm,2} - (\mathfrak{Q}_T^{\pm,1})^2}, \end{cases} \tag{4.7}$$

which can be rewritten as

$$\begin{pmatrix} \alpha_T^{\pm} \\ \beta_T^{\pm} \end{pmatrix} = \begin{pmatrix} a_{\pm} \\ b_{\pm} \end{pmatrix} + \sigma_{\pm} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathfrak{Q}_T^{\pm,2} & \mathfrak{Q}_T^{\pm,1} \\ \mathfrak{Q}_T^{\pm,1} & \mathfrak{Q}_T^{\pm,0} \end{pmatrix}^{-1} \begin{pmatrix} M_T^{\pm,1} \\ M_T^{\pm,0} \end{pmatrix}. \tag{4.8}$$

Proof. Note that $\mathfrak{Q}_T^{\pm,0}\mathfrak{Q}_T^{\pm,2} - (\mathfrak{Q}_T^{\pm,1})^2$ is $\mathbb{P}$-a.s. positive by Cauchy–Schwarz. This, and replacing the equalities (4.6) in (2.5) completes the proof. □

Proof of Item (ii) *of Theorem 1.* The proof follows from combining equation (4.7) in Lemma 3 with (Lépingle, 1995, Theorem 1, p.150) and Lemma 2.

□

Proof of Item (iii) *of Theorem 1.* The proof follows from Lemma 2, (2.6), and Theorem 2.2 in (Crimaldi and Pratelli, 2005). □

Proof of Item (iv) *of Theorem 1.* By (4.6), the following holds

$$\log \frac{G_T(a_+ + \frac{1}{\sqrt{T}}\Delta a_+, b_+ + \frac{1}{\sqrt{T}}\Delta b_+, a_- + \frac{1}{\sqrt{T}}\Delta a_-, b_- + \frac{1}{\sqrt{T}}\Delta b_-)}{G_T(a_+, b_+, a_-, b_-)}$$
$$= \sum_{\pm \in \{+,-\}} \left(\frac{1}{\sqrt{T}\sigma_{\pm}} \begin{pmatrix} \Delta a_{\pm} \\ \Delta b_{\pm} \end{pmatrix} \cdot A_T^{\pm} - \frac{1}{2T\sigma_{\pm}^2} \begin{pmatrix} \Delta a_{\pm} \\ \Delta b_{\pm} \end{pmatrix} \cdot \langle A^{\pm}, A^{\pm} \rangle_T \begin{pmatrix} \Delta a_{\pm} \\ \Delta b_{\pm} \end{pmatrix} \right),$$

where $A_T^{\pm} := \begin{pmatrix} -M_T^{\pm,1} \\ M_T^{\pm,0} \end{pmatrix}$. Note that $\langle A^{\pm}, A^{\pm} \rangle_T = \begin{pmatrix} \mathfrak{Q}_T^{\pm,2} & -\mathfrak{Q}_T^{\pm,1} \\ -\mathfrak{Q}_T^{\pm,1} & \mathfrak{Q}_T^{\pm,0} \end{pmatrix}$ and $\langle A^+, A^- \rangle_T = 0$. Lemma 2 ensures that $\frac{1}{T} \langle A^{\pm}, A^{\pm} \rangle_T \xrightarrow[T\to\infty]{\text{a.s.}} \Gamma_{\pm}$. This, and the same argument as in the proof of Theorem 1.(iii) show that

$$A_T := \frac{1}{\sqrt{T}} \begin{pmatrix} \sigma_+^{-1} A_T^+ \\ \sigma_-^{-1} A_T^- \end{pmatrix} \xrightarrow[T\to\infty]{\text{law}} \mathcal{N}(0, \Gamma)$$

and $\langle A_T, A_T \rangle_T = \frac{1}{T} \begin{pmatrix} \sigma_+^{-2} \langle A^+, A^+ \rangle_T & 0 \\ 0 & \sigma_-^{-2} \langle A^-, A^- \rangle_T \end{pmatrix} \xrightarrow[T\to\infty]{\text{a.s.}} \Gamma.$ □

4.3 Proof of Theorem 2

The proof of Item (i) of Theorem 2 is analogous to the proof of Item (i) of Theorem 1, and is therefore omitted. The proofs of Items (ii)–(iii) of Theorem 2 follow from Lemma 4. Specifically, for all $N \in \mathbb{N}$,

$$\begin{aligned} &\left(\widehat{a}_{T_N,N}^{\pm} - a_{\pm},\ \widehat{b}_{T_N,N}^{\pm} - b_{\pm} \right) \\ &= \left(\widehat{a}_{T_N,N}^{\pm} - \alpha_{T_N}^{\pm},\ \widehat{b}_{T_N,N}^{\pm} - \beta_{T_N}^{\pm} \right) + \left(\alpha_{T_N}^{\pm} - a_{\pm},\ \beta_{T_N}^{\pm} - b_{\pm} \right). \end{aligned}$$

The second term of the sum is handled using Theorem 1 (specifically, Item (iii)), providing the desired limit distribution. The first term can be rewritten using equations (2.5) and (2.10) as an expression that involves only terms of the kind

$$\left(\frac{\mathfrak{Q}_{T_N,N}^{\pm,i}}{\mathfrak{Q}_{T_N,N}^{\pm,0} \mathfrak{Q}_{T_N,N}^{\pm,2} - (\mathfrak{Q}_{T_N,N}^{\pm,1})^2} - \frac{\mathfrak{Q}_{T_N}^{\pm,i}}{\mathfrak{Q}_{T_N}^{\pm,0} \mathfrak{Q}_{T_N}^{\pm,2} - (\mathfrak{Q}_{T_N}^{\pm,1})^2} \right) \mathfrak{M}_{T_N}^{\pm,j} + \frac{\mathfrak{Q}_{T_N,N}^{\pm,i} (\mathfrak{M}_{T_N,N}^{\pm,j} - \mathfrak{M}_{T_N}^{\pm,j})}{\mathfrak{Q}_{T_N,N}^{\pm,0} \mathfrak{Q}_{T_N,N}^{\pm,2} - (\mathfrak{Q}_{T_N,N}^{\pm,1})^2},$$

for $j \in \{0,1\}$ and $i \in \{0,1,2\}$,

Combining Lemma 4 with Lemma 2 and Theorem 2.2 in (Crimaldi and Pratelli, 2005) ensures the consistency of the estimator if $\Delta_N \to 0$ as $N \to \infty$. Furthermore, if $T_N \Delta_N \to 0$ as $N \to \infty$, then it also implies that

$$\sqrt{T_N}\left(\widehat{a}^{\pm}_{T_N,N} - \alpha^{\pm}_{T_N},\ \widehat{b}^{\pm}_{T_N,N} - \beta^{\pm}_{T_N}\right) \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

Lemma 4. *Assume the process is ergodic. Let X be the solution to* (1.1)*, with X_0 distributed as the stationary distribution μ in* (4.2)*. In addition, let $\lambda \in \{1,2\}$ be fixed, and let $(T_N)_{N\in\mathbb{N}} \subset (0,\infty)$ be a sequence satisfying that, as $N \to \infty$, $T_N \to \infty$ and $\lim_{N\to\infty} T_N^{1-1/\lambda}\sqrt{\Delta_N} = 0$, where $\Delta_N := \sup_{k=1,\dots,N}(t_k - t_{k-1})$. Then, for all $m \in \{0,1,2\}$ and $j \in \{0,1\}$, the following holds:*

$$\limsup_{N\to\infty} T_N^{-1/\lambda}\mathbb{E}\left[|\mathfrak{Q}^{\pm,m}_{T_N} - \mathfrak{Q}^{\pm,m}_{T_N,N}|\right] = 0 \text{ and } \limsup_{N\to\infty} T_N^{-1/\lambda}\mathbb{E}\left[|\mathfrak{M}^{\pm,j}_{T_N} - \mathfrak{M}^{\pm,j}_{T_N,N}|\right] = 0,$$

where $\mathfrak{Q}^{\pm,m}_{T_N}$, $\mathfrak{Q}^{\pm,m}_{T_N,N}$, $\mathfrak{M}^{\pm,j}_{T_N}$, and $\mathfrak{M}^{\pm,j}_{T_N,N}$ are defined in (2.2) *and* (2.9)*.*

Proof of Item (iv) *of Theorem 2.* The proof is similarly to that of Item (iv) of Theorem 1. The analogous version of $A^{\pm}_T$ is

$$A^{N,\pm}_{T_N} := \left(-\mathfrak{M}^{\pm,1}_{T_N,N} - \mathfrak{a}_{\pm}\mathfrak{Q}^{\pm,2}_{T_N,N} + \mathfrak{b}_{\pm}\mathfrak{Q}^{\pm,1}_{T_N,N},\quad \mathfrak{M}^{\pm,0}_{T_N,N} - \mathfrak{b}_{\pm}\mathfrak{Q}^{\pm,0}_{T_N,N} + \mathfrak{a}_{\pm}\mathfrak{Q}^{\pm,1}_{T_N,N}\right),$$

and Lemma 4 ensures that the asymptotic behavior of the latter quantity is the same as that of the continuous-time analogue, given in Theorem 2.(iv).

□

Supplementary Material

The online Supplementary Material contains proofs of Lemma 4 and Theorem 3, and an extension of the results to the multi-threshold case in Section S3.

Acknowledgments

We are grateful to G. Conforti, A. Lejay, and E. Mariucci for their discussions, and to two referees for their helpful remarks and suggestions.

References

Ait-Sahalia, Y. (1996). Testing Continuous-Time Models of the Spot Interest Rate. *The Review of Financial Studies 9*(2), 385–426.

Amorino, C. and A. Gloter (2020). Contrast function estimation for the drift parameter of ergodic jump diffusion process. *Scandinavian Journal of Statistics 47*(2), 279–346.

Ang, A. and G. Bekaert (2002a). Regime Switches in Interest Rates. *Journal of Business & Economic Statistics 20*(2), 163–182.

Ang, A. and G. Bekaert (2002b). Short rate nonlinearities and regime switches. *Journal of Economic Dynamics and Control 26*(7), 1243 – 1274. Finance.

Ang, A., G. Bekaert, and M. Wei (2008). The Term Structure of Real Rates and Expected

Inflation. *The Journal of Finance 63*(2), 797–849.

Ang, A. and A. Timmermann (2012). Regime Changes and Financial Markets. *Annual Review of Financial Economics 4*, 313–337.

Archontakis, T. and W. Lemke (2008a). Bond pricing when the short-term interest rate follows a threshold process. *Quantitative Finance 8*(8), 811–822.

Archontakis, T. and W. Lemke (2008b). Threshold Dynamics of Short-term Interest Rates: Empirical Evidence and Implications for the Term Structure. *Economic Notes 37*(1), 75–117.

Ben Alaya, M. and A. Kebaier (2013). Asymptotic Behavior of the Maximum Likelihood Estimator for Ergodic and Nonergodic Square-Root Diffusions. *Stochastic Analysis and Applications 31*(4), 552–573.

Black, F. and P. Karasinski (1991). Bond and Option Pricing When Short Rates Are Lognormal. *Financial Analysts Journal 47*(4), 52–59.

Bokil, V., N. Gibson, S. Nguyen, E. Thomann, and E. Waymire (2020). An Euler-Maruyama method for diffusion equations with discontinuous coefficients and a family of interface conditions. *Journal of Computational and Applied Mathematics 368*, 112545.

Brockwell, P. J. and R. J. Williams (1997). On the Existence and Application of Continuous-Time Threshold Autoregressions of Order Two. *Advances in Applied Probability 29*(1), 205–227.

Chan, K. S. (1993, 03). Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model. *Ann. Statist. 21*(1), 520–533.

Chen, C. W. S., M. K. P. So, and F.-C. Liu (2011). A review of threshold time series models in finance. *Statistics and its Interface 4*(2), 167–181.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). An Intertemporal General Equilibrium Model of Asset Prices. *Econometrica 53*(2), 363–384.

Crimaldi, I. and L. Pratelli (2005). Convergence results for multivariate martingales. *Stochastic Process. Appl. 115*(4), 571–577.

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

Dieker, A. B. and X. Gao (2013, 08). Positive recurrence of piecewise Ornstein-Uhlenbeck processes and common quadratic Lyapunov functions. *Ann. Appl. Probab. 23*(4), 1291–1317.

Ding, K., Z. Cui, and Y. Wang (2020). A Markov chain approximation scheme for option pricing under skew diffusions. *Quantitative Finance*, 1–20.

Dong, F. and H. Y. Wong (2017, September). Variance Swaps under the Threshold Ornstein-Uhlenbeck Model. *Appl. Stoch. Model. Bus. Ind. 33*(5), 507–521.

Feng, A. (2016, June). Parameter Estimations for Skew Ornstein-Uhlenbeck Processes. *International Journal of Science and Research (IJSR)*, 1776–1781.

Gairat, A. and V. Shcherbakov (2016). Density of skew Brownian motion and its functionals with application in finance. *Mathematical Finance 26*(4), 1069–1088.

Gospodinov, N. (2005, 07). Testing For Threshold Nonlinearity in Short-Term Interest Rates. *Journal of Financial Econometrics 3*(3), 344–371.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics 42*(1), 27 – 62.

Hu, Y. and Y. Xi (2022). Parameter estimation for threshold Ornstein-Uhlenbeck processes from discrete observations. *Journal of Computational and Applied Mathematics 411*, 114264.

Hull, J. and A. White (1990). Pricing Interest-Rate-Derivative Securities. *Review of Financial Studies 3*(4), 573–92.

Jacod, J. and P. Protter (2012). *Discretization of processes*, Volume 67 of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg.

Jacod, J. and A. N. Shiryaev (2003). *Limit theorems for stochastic processes* (Second ed.), Volume 288 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin.

Kalimipalli, M. and R. Susmel (2004). Regime-switching stochastic volatility and short-term interest rates. *Journal of Empirical Finance 11*(3), 309 – 329.

Kessler, M. (1997). Estimation of an Ergodic Diffusion from Discrete Observations. *Scandinavian Journal of Statistics 24*(2), 211–229.

Krugman, P. R. (1991). Target zones and exchange rate dynamics. *The Quarterly Journal of*

Economics 106(3), 669–682.

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

Le Cam, L. and G. L. Yang (2000). *Asymptotics in statistics* (Second ed.). Springer Series in Statistics. Springer-Verlag, New York.

Le Gall, J.-F. (1985). One-dimensional stochastic differential equations involving the local times of the unknown process. *Stochastic Analysis. Lecture Notes Math. 1095*, 51–82.

Lejay, A. and P. Pigato (2018). Statistical estimation of the Oscillating Brownian Motion. *Bernoulli 24*(4B), 3568–3602.

Lejay, A. and P. Pigato (2019). 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).

Lejay, A. and P. Pigato (2020). Maximum likelihood drift estimation for a threshold diffusion. *Scandinavian Journal of Statistics 47*(3), 609–637.

Lépingle, D. (1995). Euler scheme for reflected stochastic differential equations. *Math. Comput. Simulation 38*(1-3), 119–126. Probabilités numériques (Paris, 1992).

Lipton, A. (2018). Oscillating Bachelier and Black-Scholes Formulas. In *Financial Engineering*. World Scientific.

Lipton, A. and A. Sepp (2011). Filling the gaps. *Risk Magazine*, 66–71.

Mazzonetto, S. (2019). Rates of convergence to the local time of Oscillating and Skew Brownian Motions. *arXiv preprint arXiv:1912.04858*.

Pai, J. and H. Pedersen (1999). Threshold Models of the Term Structure of Interest Rate. In *Joint day Proceedings Volume of the XXXth International ASTIN Colloquium/9th International AFIR Colloquium, Tokyo, Japan*, pp. 387–400.

Pfann, G. A., P. C. Schotman, and R. Tschernig (1996). Nonlinear interest rate dynamics and implications for the term structure. *Journal of Econometrics 74*(1), 149 – 176.

Pigato, P. (2019). Extreme at-the-money skew in a local volatility model. *Finance and Stochastics 23*, 827–859.

Rabemananjara, R. and J. M. Zakoian (1993, January). Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics 8*(1), 31–49.

Rényi, A. (1963). On stable sequences of events. *Sankhyā Ser. A 25*, 293–302.

Su, F. and K.-S. Chan (2015). Quasi-likelihood estimation of a threshold diffusion process. *J. Econometrics 189*(2), 473–484.

Su, F. and K.-S. Chan (2017). Testing for threshold diffusion. *J. Bus. Econom. Statist. 35*(2), 218–227.

Tong, H. (1983). *Threshold models in nonlinear time series analysis*, Volume 21 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University

Press.

Tong, H. (2011). Threshold models in time series analysis — 30 years on. *Statistics and its Interface 4*.

Tong, H. (2015). Threshold models in time series analysis—some reflections. *J. Econometrics 189*(2), 485–491.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*(2), 177 – 188.

Wu, Y. and H. Zhang (1996). Mean Reversion in Interest Rates: New Evidence from a Panel of OECD Countries. *Journal of Money, Credit and Banking 28*(4), 604–621.

Xing, X., D. Zhao, and B. Li (2020). Parameter estimation for the skew Ornstein-Uhlenbeck processes based on discrete observations. *Communications in Statistics - Theory and Methods 49*(9), 2176–2188.

Yadav, P., P. Pope, and K. Paudyal (1994). Threshold autoregressive modeling in finance: The price differences of equivalent assets. *Mathematical Finance 4*(2), 205–221.

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

Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France

E-mail: sara.mazzonetto@univ-lorraine.fr

Department of Economics and Finance, University of Rome Tor Vergata, Italy

E-mail: paolo.pigato@uniroma2.it