Ann Inst Stat Math (2012) 64:383–413
DOI 10.1007/s10463-010-0318-1

On identification of the threshold diffusion processes

Yury A. Kutoyants

Received: 14 December 2009 / Revised: 28 June 2010 / Published online: 17 November 2010

Abstract We consider the problems of parameter estimation for several models of threshold ergodic diffusion processes in the asymptotics of large samples. These models are the direct continuous time analogues of the well known in time series analysis threshold autoregressive models. In such models, the trend is switching when the observed process attaints some (unknown) values and the problem is to estimate it or to test some hypotheses concerning these values. The related statistical problems correspond to the singular estimation or testing, for example, the rate of convergence of estimators is T and not $\sqrt{T}$ as in regular estimation problems. We study the asymptotic behavior of the maximum likelihood and Bayesian estimators and discuss the possibility of the construction of the goodness-of-fit test for such models of observation.

Keywords Parameter estimation · Threshold models · Singular estimation · Ergodic diffusion process · Goodness-of-fit test · Cramer-von Mises type tests

1 Introduction

The simplest example of the threshold model is the following threshold autoregressive (TAR) time series:

$$X_{j+1} = \varrho_1 X_j \, \mathbb{1}_{\{X_j < \vartheta\}} + \varrho_2 X_j \, \mathbb{1}_{\{X_j \geq \vartheta\}} + \varepsilon_{j+1}, \quad j = 0, \ldots, n-1, \tag{1}$$

Y. A. Kutoyants (✉)
Laboratoire de Statistique et Processus,
Université du Maine, av. O. Messiaen,
72085 Le Mans Cédex 9, France
e-mail: kutoyants@univ-lemans.fr

where ε_j are i.i.d. $\mathcal{N}(0, s^2)$, $\varrho_1 \neq \varrho_2$ and $|\varrho_i| < 1$. Therefore, we have two different autoregressive processes depending on the region of observations $\{x : x < \vartheta\}$ or $\{x : x \geq \vartheta\}$. This time series has ergodic properties with invariant density close to a weighted sum of two Gaussian densities. If we suppose that $s^2, \varrho_1, \varrho_2$ are known and $\vartheta \in \Theta = (\alpha, \beta)$ is unknown parameter, then we obtain the first problem of threshold ϑ estimation. It is easy to see that the likelihood ratio is a piece-wise constant (discontinuous) function of ϑ, the Fisher information is equal infinity (see e.g., Chan and Kutoyants 2008). As usual in singular estimation problems, the rate of convergence of maximum likelihood $\hat{\vartheta}_n$ or Bayesian $\tilde{\vartheta}_n$ estimators is n and not $\sqrt{n}$ i.e.; the quantities $n(\hat{\vartheta}_n - \vartheta)$ and $n(\tilde{\vartheta}_n - \vartheta)$ have non-degenerate limits.

There are many different threshold regression models of such type extensively developed in econometrics and, of course, the identification of these models attracts attention of statisticians (see e.g. the works by Quandt 1958; Tong 1990; Chan 1993; Hansen 2000; Fan and Yao 2003; Koul et al. 2003; Chan and Kutoyants 2008 and the references therein). Note that continuous time models actually find a wide range of applications in econometrical problems and occupy a central place in financial mathematics (see e.g., the work by Shreve 2004).

Our goal is to study several models of continuous time analogues (diffusion processes) of such threshold type time series and to describe the properties of estimators of the thresholds for these models. Note that the general theory of parameter estimation (in regular case) for ergodic diffusion processes is actually well developed (see e.g. Kutoyants 2004; Yoshida 2009 and references therein), but the problems of threshold estimation are of singular type and need a special consideration. To illustrate these statements of the problem, let us consider the following process

$$\mathrm{d}X_t = -\rho_1 X_t \mathbb{1}_{\{X_t<\vartheta\}}\mathrm{d}t - \rho_2 X_t \mathbb{1}_{\{X_t\geq\vartheta\}}\mathrm{d}t + \sigma \mathrm{d}W_t, \quad 0 \leq t \leq T, \tag{2}$$

where W_t is Wiener process, $\rho_1 \neq \rho_2$ and $\rho_i > 0$. We call it *Threshold Ornstein–Uhlenbeck* (TOU) process because it can be considered as a mixture of two different Ornstein–Uhlenbeck processes with switching. If we suppose that σ, ρ_1, ρ_2 are known and $\vartheta \in \Theta = (\alpha, \beta)$ is unknown parameter then we obtain the problem of parameter (threshold) ϑ estimation.

It is in some sense similar to TAR (1) and the link between them can be clarified by the following consideration. Let us consider the discrete time approximation of the process (2) with $t_j = j\delta$, $j = 1, \ldots, n-1$, where $\delta = T/n$, then we obtain

$$X_{t_{j+1}} = (1-\rho_1\delta)\, X_{t_j}\, \mathbb{1}_{\left\{X_{t_j}<\vartheta\right\}} + (1-\rho_2\delta)\, X_{t_j}\, \mathbb{1}_{\left\{X_{t_j}\geq\vartheta\right\}} + \sigma\, \left[W_{t_{j+1}} - W_{t_j}\right].$$

This process coincides with (1) if we put $X_j = X_{t_j}, \varrho_i = (1-\rho_i\,\delta)$ and $\varepsilon_{j+1} = \sigma\,\left[W_{t_{j+1}} - W_{t_j}\right] \sim \mathcal{N}(0, \sigma^2\delta)$, i.e., $s^2 = \sigma^2\delta$. Hence, the regression model (1) is a discrete time approximation of the TOU process (2). Note that these continuous time threshold processes are studied by Decamps et al. (2006) as interest rates models.

The threshold estimation problems for both models are of singular type and the limit distributions of the MLE's $n(\hat{\vartheta}_n - \vartheta)$ and $T(\hat{\vartheta}_T - \vartheta)$ are of *argsup* type functionals of the compound Poisson and Wiener processes, respectively.

The process $(X_t)_{t\geq 0}$ has ergodic properties, the invariant density is a mixture of two Gaussian, the Fisher information is equal to infinity and we show that the maximum likelihood and Bayesian estimators converge to two different limit laws.

We consider several other threshold type models of ergodic diffusion processes and study the asymptotic properties of the maximum likelihood and Bayesian estimators. The main result of this work is Theorem 1 in Sect. 3, where we describe the properties of estimators in the case of nonlinear trend coefficients and many thresholds exist.

We discuss as well the construction of the goodness-of-fit tests for such threshold models.

2 Threshold Ornstein–Uhlenbeck process

2.1 Threshold estimation

We start with the TOU process

$$\mathrm{d}X_t = -\rho_1 X_t\, \mathbb{1}_{\{X_t<\vartheta\}}\mathrm{d}t - \rho_2 X_t\, \mathbb{1}_{\{X_t\geq\vartheta\}}\mathrm{d}t + \sigma \mathrm{d}W_t, \quad X_0, \quad 0 \leq t \leq T, \tag{3}$$

where we suppose that the following condition is fulfilled.

Condition $\mathcal{A}_*$. The constants $\rho_1 \neq \rho_2$, $\rho_i > 0$ and $\sigma^2 > 0$ are known and the parameter $\vartheta \in \Theta = (\alpha, \beta), \alpha > 0$ is unknown. The initial value X_0 is independent on the Wiener process random variable.

The value $\vartheta = 0$ is excluded because in the case $\vartheta = 0$ there is no jump in the trend coefficient and the properties of estimators are quite different. This is due to the property: $-\rho_1 x = -\rho_2 x = 0$ if $x = 0$ and therefore there is no switching at this point. As $\rho_1 \neq \rho_2$ the MLE is consistent and asymptotically normal with “regular rate” $\sqrt{T}$.

We consider the problem of estimation of the threshold ϑ by the continuous time observations $X^T = (X_t, 0 \leq t \leq T)$ and we are interested in the asymptotic behavior of estimators as $T \to \infty$.

Note that the conditions $\mathcal{ES}$ of the existence of solution and $\mathcal{RP}$ of the ergodicity are fulfilled (see Kutoyants 2004, Sects. 1.1, 1.2) and the process $(X_t)_{t\geq 0}$ has ergodic properties with the invariant density

$$f(\vartheta, x) = p_1(x, \vartheta)\; e^{-\frac{\rho_1 (x^2-\vartheta^2)}{\sigma^2}} + p_2(x, \vartheta)\, e^{-\frac{\rho_2 (x^2-\vartheta^2)}{\sigma^2}}.$$

Here $p_1(x, \vartheta) = G(\vartheta)^{-1}\, \mathbb{1}_{\{x<\vartheta\}}$, and $p_2(x, \vartheta) = G(\vartheta)^{-1}\, \mathbb{1}_{\{x\geq\vartheta\}}$ and $G(\vartheta)$ is the normalizing constant. To simplify the exposition we suppose that the random variable X_0 has the density function $f(\vartheta, x)$, hence the observed process is stationary.

We are interested in the asymptotic behavior of the maximum likelihood estimator (MLE) and Bayesian estimator (BE) of the parameter ϑ; therefore, we need the likelihood ratio function $L\left(\vartheta, X^T\right)$. This function can be written as (see Liptser and Shiryayev 2001)

$$\ln L\left(\vartheta, X^T\right) = -\frac{\rho_1}{\sigma^2}\int_0^T X_t \mathbb{1}_{\{X_t<\vartheta\}}\,\mathrm{d}X_t - \frac{\rho_2}{\sigma^2}\int_0^T X_t \mathbb{1}_{\{X_t\geq\vartheta\}}\,\mathrm{d}X_t$$
$$-\frac{\rho_1^2}{2\sigma^2}\int_0^T X_t^2 \mathbb{1}_{\{X_t<\vartheta\}}\,\mathrm{d}t - \frac{\rho_2^2}{2\sigma^2}\int_0^T X_t^2 \mathbb{1}_{\{X_t\geq\vartheta\}}\,\mathrm{d}t + \ln f\left(\vartheta, X_0\right).$$

The contribution of the term $\ln f\left(\vartheta, X_0\right)$ is asymptotically negligeable and we will always omitted it for simplicity of exposition (see the details in Kutoyants 2004).

The MLE $\hat{\vartheta}_T$ and BE (for quadratic loss function) $\tilde{\vartheta}_T$ are defined as usual by the relations

$$L\left(\hat{\vartheta}_T, X^T\right) = \sup_{\theta\in\Theta} L\left(\theta, X^T\right) \quad \text{and} \quad \tilde{\vartheta}_T = \frac{\int_\alpha^\beta \theta\, p\left(\theta\right) L\left(\theta, X^T\right)\mathrm{d}\theta}{\int_\alpha^\beta p\left(\theta\right) L\left(\theta, X^T\right)\mathrm{d}\theta}. \tag{4}$$

Here $p\left(\theta\right), \alpha \leq \theta \leq \beta$ is positive on Θ continuous density a priori.

To describe theirs properties, we need the following notations. Let us introduce

– the random process

$$Z_0\left(u\right) = \exp\left\{W\left(u\right) - \frac{|u|}{2}\right\}, \quad u \in \mathscr{R},$$

where $W\left(\cdot\right)$ is two-sided Wiener process,

– two random variables $\hat{u}$ and $\tilde{u}$ defined by the relations

$$Z_0\left(\hat{u}\right) = \sup_{u\in\mathscr{R}} Z_0\left(u\right), \qquad \tilde{u} = \frac{\int_{\mathscr{R}} u\, Z_0\left(u\right)\ \mathrm{d}u}{\int_{\mathscr{R}} Z_0\left(u\right)\ \mathrm{d}u} \tag{5}$$

– the function

$$\Gamma_\vartheta^2 = \frac{\left(\rho_2-\rho_1\right)^2\ \vartheta^2}{G\left(\vartheta\right)\ \sigma^2}\, e^{-\frac{\rho_1^2\vartheta^2}{\sigma^2}}.$$

The function Γ_ϑ^2 is equal to the square of the jump divided by the diffusion coefficient at the point ϑ and multiplied by the value of the invariant density at the point of jump (see e.g. (16) below).

The properties of estimators are given in the following proposition.

Proposition 1 *Let the condition $\mathcal{A}_*$ be fulfilled, then the MLE $\hat{\vartheta}_T$ and the BE $\tilde{\vartheta}_T$ are uniformly on compacts $\mathbb{K} \subset \Theta$ consistent: for any $\nu > 0$*

$$\sup_{\vartheta\in\mathbb{K}} \mathbf{P}_\vartheta\left\{\left|\hat{\vartheta}_T - \vartheta\right| > \nu\right\} \longrightarrow 0,$$

have two different limit distributions

$$T\left(\hat{\vartheta}_T-\vartheta\right)\Longrightarrow\frac{\hat{u}}{\Gamma_\vartheta^2},\quad T\left(\tilde{\vartheta}_T-\vartheta\right)\Longrightarrow\frac{\tilde{u}}{\Gamma_\vartheta^2},$$

theirs moments converge: for any $p>0$

$$\mathbf{E}_\vartheta\left|T\left(\hat{\vartheta}_T-\vartheta\right)\right|^p\longrightarrow\mathbf{E}\left|\frac{\hat{u}}{\Gamma_\vartheta^2}\right|^p,\quad \mathbf{E}_\vartheta\left|T\left(\tilde{\vartheta}_T-\vartheta\right)\right|^p\longrightarrow\mathbf{E}\left|\frac{\tilde{u}}{\Gamma_\vartheta^2}\right|^p.$$

For the proof, see Sect. 5.

Note that the same normalization and the same type limits (with different Γ_ϑ), we have in the problem of delay ϑ estimation by the observations of the following Gaussian process

$$\mathrm{d}X_t=-\rho\,X_{t-\vartheta}\,\mathrm{d}t+\sigma\mathrm{d}W_t,\quad 0\le t\le T$$

see details in Küchler and Kutoyants (2000) (or in Kutoyants 2004, Sect. 3.3).

Remind that the BEs are usually asymptotically efficient in singular parameter estimation problems (Ibragimov and Khasminskii 1981). The following lower bound is valid: for all estimators $\bar{\vartheta}_T$

$$\underline{\lim}_{\delta\to 0}\,\underline{\lim}_{T\to\infty}\,\sup_{|\vartheta-\vartheta_0|<\delta}T^2\,\mathbf{E}_\vartheta\left(\bar{\vartheta}_T-\vartheta\right)^2\ge\frac{\mathbf{E}\tilde{u}^2}{\Gamma_{\vartheta_0}^4}$$

see Ibragimov and Khasminskii (1981), Sect. 1.9 (or Kutoyants 2004, Proposition 2.24). We call an estimator ϑ_T^* asymptotically efficient if for all $\vartheta_0\in\Theta$ we have the equality

$$\lim_{\delta\to 0}\,\lim_{T\to\infty}\,\sup_{|\vartheta-\vartheta_0|<\delta}T^2\,\mathbf{E}_\vartheta\left(\vartheta_T^*-\vartheta\right)^2=\frac{\mathbf{E}\tilde{u}^2}{\Gamma_{\vartheta_0}^4}.$$

It can be verified that the convergence of the moments of BEs is uniform on the compacts in Θ and that the function Γ_ϑ is continuous. From these properties, we obtain immediately the asymptotic efficiency of the BEs (in the sense of this lower bound).

The quantities $\mathbf{E}\hat{u}^2$ and $\mathbf{E}\tilde{u}^2$ were calculated by

$$\mathbf{E}\hat{u}^2=26\quad>\quad\mathbf{E}\tilde{u}^2=16\zeta\left(3\right)\sim 19.2$$

where $\zeta\left(\cdot\right)$ is Riemann zeta function. This relation shows the difference between the limit variances of the MLE and BE.

2.2 All parameters unknown

It is possible to describe the properties of estimators in the case when all three parameters $(\rho_1,\rho_2,\vartheta)=(\vartheta_1,\vartheta_2,\vartheta_3)=\boldsymbol{\vartheta}\in\boldsymbol{\Theta}$ are unknown and we observe

$$\mathrm{d}X_t=-\vartheta_1 X_t\,\mathbb{1}_{\{X_t<\vartheta_3\}}\mathrm{d}t-\vartheta_2 X_t\,\mathbb{1}_{\{X_t\geq\vartheta_3\}}\mathrm{d}t+\sigma\mathrm{d}W_t,\quad 0\leq t\leq T. \tag{6}$$

We have $\boldsymbol{\Theta}=(\alpha_1,\beta_1)\times(\alpha_2,\beta_2)\times(\alpha_3,\beta_3)$. Let us denote by ξ the random variable with the density $f(\boldsymbol{\vartheta},x)$.

Proposition 2 *Suppose that $\beta_1<\alpha_2$ and $\alpha_2>0$, then the MLE $\hat{\boldsymbol{\vartheta}}_T$, BE $\tilde{\boldsymbol{\vartheta}}_T$ are consistent, have the following limit distributions*

$$\sqrt{T}\left(\hat\vartheta_{1,T}-\vartheta_1\right)\Longrightarrow\zeta_1\sim\mathcal{N}\left(0,\frac{\sigma^2}{\mathbf{E}_{\boldsymbol{\vartheta}}\ \xi^2\mathbb{1}_{\{\xi<\vartheta_3\}}}\right),$$
$$\sqrt{T}\left(\hat\vartheta_{2,T}-\vartheta_2\right)\Longrightarrow\zeta_2\sim\mathcal{N}\left(0,\frac{\sigma^2}{\mathbf{E}_{\boldsymbol{\vartheta}}\ \xi^2\mathbb{1}_{\{\xi\geq\vartheta_3\}}}\right),$$
$$T\left(\hat\vartheta_{3,T}-\vartheta_3\right)\Longrightarrow\frac{\hat u}{\Gamma_{\boldsymbol{\vartheta}}^2},\qquad T\left(\tilde\vartheta_{3,T}-\vartheta_3\right)\Longrightarrow\frac{\tilde u}{\Gamma_{\boldsymbol{\vartheta}}^2}.$$

The BE $\tilde\vartheta_{1,T},\tilde\vartheta_{2,T}$ *have the same asymptotic properties as $\hat\vartheta_{1,T},\hat\vartheta_{2,T}$, the random variables ζ_1 and ζ_2 are independent and are independent of $\hat u$, $\tilde u$.*

See the proof in Sect. 5.

The construction of the MLE can be slightly simplified by the following “separation”.

The MLE of the first two components can be written as

$$\hat\vartheta_{1,T}=-\frac{\int_0^T X_t\,\mathbb{1}_{\left\{X_t<\hat\vartheta_{3,T}\right\}}\mathrm{d}X_t}{\int_0^T X_t^2\,\mathbb{1}_{\left\{X_t<\hat\vartheta_{3,T}\right\}}\mathrm{d}t},\qquad \hat\vartheta_{2,T}=-\frac{\int_0^T X_t\,\mathbb{1}_{\left\{X_t\geq\hat\vartheta_{3,T}\right\}}\mathrm{d}X_t}{\int_0^T X_t^2\,\mathbb{1}_{\left\{X_t\geq\hat\vartheta_{3,T}\right\}}\mathrm{d}t}$$

but to study these expressions can be quite difficult because the estimator $\hat\vartheta_{3,T}$ depends on the whole trajectory X^T and therefore the random function $X_t\,\mathbb{1}_{\{X_t<\hat\vartheta_{3,T}\}},\ 0\leq t\leq T$ depends of the “future”. Hence, the stochastic integral needs a special treatment. The problem can be simplified as follows: let us estimate the parameter ϑ_3 by the first $X^{\sqrt{T}}=\{X_t,0\leq t\leq\sqrt{T}\}$ observation and denote by $\vartheta^*_{3,\sqrt{T}}$ the corresponding consistent estimator. We suppose that there exists $b>0$ such that

$$\mathbf{P}_{\boldsymbol{\vartheta}}\left\{\left|\vartheta^*_{3,\sqrt{T}}-\vartheta_3\right|>T^{-b}\right\}\longrightarrow 0 \tag{7}$$

as $T \to \infty$. Then, we define the estimators (see the comment at the end of Sect. 2)

$$\vartheta_{1,T}^{\circ} = -\frac{\int_{\sqrt{T}}^{T} X_t \, \mathbb{1}_{\left\{X_t < \vartheta_{3,\sqrt{T}}^*\right\}} \, \mathrm{d}X_t}{\int_{\sqrt{T}}^{T} X_t^2 \, \mathbb{1}_{\left\{X_t < \vartheta_{3,\sqrt{T}}^*\right\}} \, \mathrm{d}t}, \quad \vartheta_{2,T}^{\circ} = -\frac{\int_{\sqrt{T}}^{T} X_t \, \mathbb{1}_{\left\{X_t \geq \vartheta_{3,\sqrt{T}}^*\right\}} \, \mathrm{d}X_t}{\int_{\sqrt{T}}^{T} X_t^2 \, \mathbb{1}_{\left\{X_t \geq \vartheta_{3,\sqrt{T}}^*\right\}} \, \mathrm{d}t}. \qquad (8)$$

Now the stochastic integrals are well defined and the consistency and asymptotic normality of these estimators follow from the usual limit theorems, i.e., we have

$$\sqrt{T}\left(\vartheta_{1,T}^{\circ} - \vartheta_1\right) = -\sigma \frac{\frac{1}{\sqrt{T}} \int_{\sqrt{T}}^{T} X_t \, \mathbb{1}_{\left\{X_t < \vartheta_{3,\sqrt{T}}^*\right\}} \, \mathrm{d}W_t}{\frac{1}{T} \int_{\sqrt{T}}^{T} X_t^2 \, \mathbb{1}_{\left\{X_t < \vartheta_{3,\sqrt{T}}^*\right\}} \, \mathrm{d}t}$$

with (law of large numbers; see Sect. 5)

$$\frac{1}{T} \int_{\sqrt{T}}^{T} X_t^2 \, \mathbb{1}_{\left\{X_t < \vartheta_{3,\sqrt{T}}^*\right\}} \, \mathrm{d}t \longrightarrow \mathbf{E}_\theta \, \xi^2 \mathbb{1}_{\{\xi < \vartheta_3\}} \qquad (9)$$

and (central limit theorem)

$$\frac{1}{\sqrt{T}} \int_{\sqrt{T}}^{T} X_t \, \mathbb{1}_{\left\{X_t < \vartheta_{3,\sqrt{T}}^*\right\}} \, \mathrm{d}W_t \Longrightarrow \zeta \sim \mathcal{N}\left(0, \mathbf{E}_\theta \, \xi^2 \mathbb{1}_{\{\xi < \vartheta_3\}}\right).$$

hence

$$\sqrt{T}\left(\vartheta_{1,T}^{\circ} - \vartheta_1\right) \Longrightarrow \mathcal{N}\left(0, \frac{\sigma^2}{\mathbf{E}_\theta \, \xi^2 \mathbb{1}_{\{\xi < \vartheta_3\}}}\right).$$

Note that the independence of the random variables ζ_1 and ζ_2 follows from the following property of stochastic integral

$$\mathbf{E}_\theta \left(\int_0^T X_t \, \mathbb{1}_{\{X_t < \vartheta_3\}} \, \mathrm{d}W_t \int_0^T X_t \, \mathbb{1}_{\{X_t \geq \vartheta_3\}} \, \mathrm{d}W_t \right) = 0.$$

The possibility to simplify the estimation of ϑ_3 we discuss at the end of the next section.

2.3 Misspecification

Let us return to the initial problem of threshold estimation and suppose that the observed process is

$$\mathrm{d}X_t = -\rho_1 X_t \, \mathbb{1}_{\{X_t < \vartheta_0\}} \mathrm{d}t - \rho_2 X_t \, \mathbb{1}_{\{X_t \geq \vartheta_0\}} \mathrm{d}t + h\left(X_t\right) \mathrm{d}t + \sigma \mathrm{d}W_t, \qquad (10)$$

where $h(\cdot)$ is some unknown function (contamination) and ϑ_0 is the true value. We assume that the statistician uses this model without $h(\cdot)$ (wrong model) and tries to estimate ϑ, i.e., he (or she) supposes that the observed process is TOU (3) and construct, say, the MLE $\hat{\vartheta}_T$ as if $h(\cdot) \equiv 0$. Then, he substitutes the observations (10) (of course, containing $h(\cdot)$). Such situation can be considered as typical for many applied problems, when there is a difference between the theoretical model and the real data. Remind that in regular case the MLE and BE are usually not consistent and converge to the value which minimizes the Kullback–Leibler distance (see Kutoyants 2004, Sect. 2.6.1). The Kullback–Leibler distance in our problem is (suppose for instant that $\vartheta_0 < \vartheta$)

$$
\begin{aligned}
D_{K-L}(\vartheta, \vartheta_0) &= \mathbf{E}^*_{\vartheta_0} \ln \frac{\mathrm{d}\mathbf{P}^*_{\vartheta_0}}{\mathrm{d}\mathbf{P}_\vartheta}\left(X^T\right) \\
&= \frac{T}{2\sigma^2}\mathbf{E}^*_{\vartheta_0}\left[\rho_1\xi\,\left[\mathbb{1}_{\{\xi<\vartheta\}} - \mathbb{1}_{\{\xi<\vartheta_0\}}\right] + \rho_2\xi\,\left[\mathbb{1}_{\{\xi\geq\vartheta\}} - \mathbb{1}_{\{\xi\geq\vartheta_0\}}\right] + h(\xi)\right]^2 \\
&= \frac{T}{2\sigma^2}\mathbf{E}^*_{\vartheta_0}\left[(\rho_1-\rho_2)\,\xi\,\mathbb{1}_{\{\vartheta_0<\xi<\vartheta\}} + h(\xi)\right]^2,
\end{aligned}
$$

where $\mathbf{E}^*_{\vartheta_0}$ denotes the expectation with respect to the measure $\mathbf{P}^*_{\vartheta_0}$ which corresponds to the process (10) (we denote its density as $f_h(\vartheta_0, x)$). It can be shown (see Kutoyants 2004, Sect. 2.6.1) that

$$
\hat{\vartheta}_T \longrightarrow \vartheta_* = \arg\inf_{\vartheta\in\Theta} D_{K-L}(\vartheta, \vartheta_0).
$$

We are interested in the following question: when $\vartheta_* = \vartheta_0$, i.e., when the MLE is nevertheless consistent? Surprisingly, it is possible even for not too small functions $h(\cdot)$. Suppose, for simplicity, that $\vartheta \in \Theta = (\alpha, \beta)$, $\alpha > 0$.

Let us introduce the function

$$
K(\vartheta, \vartheta_0) = \begin{cases} \mathbf{E}^*_{\vartheta_0}\left[(\rho_1-\rho_2)\,\xi\,\mathbb{1}_{\{\vartheta_0<\xi<\vartheta\}} + h(\xi)\right]^2, & \text{if } \vartheta \geq \vartheta_0 \\ \mathbf{E}^*_{\vartheta_0}\left[(\rho_2-\rho_1)\,\xi\,\mathbb{1}_{\{\vartheta<\xi<\vartheta_0\}} + h(\xi)\right]^2, & \text{if } \vartheta \leq \vartheta_0 \end{cases}
$$

and suppose that $\rho_2 > \rho_1$. Then for $\vartheta > \vartheta_0$ we have

$$
\begin{aligned}
K(\vartheta, \vartheta_0) &= \left[\int_{-\infty}^{\vartheta_0} + \int_{\vartheta}^{\infty}\right] h(x)^2\, f_h(\vartheta_0, x)\;\mathrm{d}x \\
&\quad + \int_{\vartheta_0}^{\vartheta} \left[(\rho_1-\rho_2)\,x + h(x)\right]^2 f_h(\vartheta_0, x)\;\mathrm{d}x,
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial K(\vartheta, \vartheta_0)}{\partial\vartheta} &= -h(\vartheta)^2\, f_h(\vartheta_0, \vartheta) + \left[(\rho_1-\rho_2)\,\vartheta + h(\vartheta)\right]^2 f_h(\vartheta_0, \vartheta) \\
&= \left[(\rho_1-\rho_2)^2\,\vartheta^2 + 2\,(\rho_1-\rho_2)\,\vartheta\, h(\vartheta)\right] f_h(\vartheta_0, \vartheta).
\end{aligned}
$$

Therefore, if

$$h(y) < \frac{y}{2}(\rho_2 - \rho_1), \quad \text{for } \alpha < y < \beta,$$

then for $\vartheta > \vartheta_0$

$$\frac{\partial K(\vartheta, \vartheta_0)}{\partial \vartheta} > 0$$

and similarly, if

$$h(y) > -\frac{y}{2}(\rho_2 - \rho_1), \quad \text{for } \alpha < y < \beta,$$

then for $\vartheta < \vartheta_0$

$$\frac{\partial K(\vartheta, \vartheta_0)}{\partial \vartheta} < 0.$$

We see that if the function $h(\cdot)$ satisfies the condition

$$|h(y)| < \frac{y}{2}(\rho_2 - \rho_1), \qquad \alpha < y < \beta, \tag{11}$$

then $\vartheta_* = \vartheta_0$ and the MLE $\hat{\vartheta}_T$ is consistent even for this "wrong model" (see Kutoyants 2004, Sect. 3.4.5 for another example). Note, that there is no conditions on $h(y)$ for $y \notin [\alpha, \beta]$.

Let us return to the problem of the construction of the preliminary consistent estimator of the parameter ϑ_3 by observations (6). Suppose that $\beta_1 - \alpha_1 < \alpha_2 - \beta_1$ and $\beta_2 - \alpha_2 < \alpha_2 - \beta_1$. Let us put

$$\hat{\vartheta}_1 = \frac{\alpha_1 + \beta_1}{2}, \qquad \hat{\vartheta}_2 = \frac{\alpha_2 + \beta_2}{2}$$

and consider the problem of estimation ϑ_3 by the "wrong model"

$$\mathrm{d}X_t = -\hat{\vartheta}_1 X_t \mathbb{1}_{\{X_t < \vartheta_3\}} \mathrm{d}t - \hat{\vartheta}_2 X_t \mathbb{1}_{\{X_t \geq \vartheta_3\}} \mathrm{d}t + \sigma \mathrm{d}W_t, \quad 0 \leq t \leq \sqrt{T}$$

with "known" $\hat{\vartheta}_1, \hat{\vartheta}_2$. This corresponds well to the model (10) with

$$h(x) = (\hat{\vartheta}_1 - \vartheta_1) x \mathbb{1}_{\{x < \vartheta_3\}} + (\hat{\vartheta}_2 - \vartheta_2) x \mathbb{1}_{\{x \geq \vartheta_3\}}.$$

We see that the condition (11) is fulfilled, hence the MLE $\hat{\vartheta}_{3,\sqrt{T}}$ is consistent and can be used in the construction of the estimators (8). Note that the estimator $\hat{\vartheta}_{3,\sqrt{T}}$ even has "singular" rate of convergence, but its limit distribution is different of that of the true MLE. Note that the similar problem was considered in Kutoyants (2004), Sect. 3.4.5.

3 General threshold model

Suppose that the observed diffusion process $X^T = \{X_t, 0 \le t \le T\}$ satisfies the equation

$$\mathrm{d}X_t = \sum_{j=1}^{k+1} S_j\left(X_t\right) \mathbb{1}_{\{\vartheta_{j-1}<X_t\le\vartheta_j\}}\mathrm{d}t + \sigma\left(X_t\right)\mathrm{d}W_t, \quad X_0, \tag{12}$$

where $\vartheta_0 = -\infty$, $\vartheta_j \in \Theta_j = \left(\alpha_j, \beta_j\right)$, $j = 1, \ldots, k$, $\vartheta_{k+1} = \infty$, $\beta_j < \alpha_{j+1}$. The unknown parameter is $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_k) \in \boldsymbol{\Theta} = \Theta_1 \times \cdots \times \Theta_k$. Our goal is to estimate $\boldsymbol{\vartheta}$ and to describe the asymptotic properties of estimators as $T \to \infty$. As before, we are interested in the estimators obtained by the maximum likelihood and Bayesian methods.

This model can be called "Nonlinear Threshold Diffusion Process". Of course, all considered above models are nonlinear due to the indicator functions. Here, we use the term "nonlinear" because the linear function ρx in the trend coefficient $-\rho x\, \mathbb{1}_{\{\cdot\}}$ is replaced by more general function $S\left(x\right)$.

$\mathcal{ES}$. The functions $S_j(\cdot)$ are locally bounded, the function $\sigma\left(\cdot\right)^2$ is continuous and positive and for some $A > 0$ the condition

$$x S_1\left(x\right) \mathbb{1}_{\{x<\alpha_1\}} + x S_{k+1}\left(x\right) \mathbb{1}_{\{x\ge\beta_k\}} + \sigma\left(x\right)^2 \le A\left(1 + x^2\right) \tag{13}$$

holds.

This condition provides the existence of unique weak solution (see Durret 1996).

The asymptotic behavior of the diffusion process is defined by the following condition.

$\mathcal{A}$. The functions $S_1\left(x\right)$, $S_{k+1}(x)$ and $\sigma\left(x\right)$ satisfy the conditions

$$\left|\sigma\left(x\right)\right|^{-1} \le B\left(1 + |x|^m\right)$$

with some $B > 0$ and $m > 0$ and

$$\varliminf_{x\to-\infty} \frac{S_1\left(x\right)}{\sigma\left(x\right)^2} > 0, \qquad \varlimsup_{x\to\infty} \frac{S_{k+1}\left(x\right)}{\sigma\left(x\right)^2} < 0.$$

By this condition the process $\left(X_t\right)_{t\ge0}$ has ergodic properties. Let us denote by $f\left(\boldsymbol{\vartheta}, x\right)$ the density of its invariant law and by ξ the random variable with such density function. Note that by this condition ξ has all polynomial moments (see Kutoyants 2004).

The *identifiability* condition in this statistical problem is the following one

$$\inf_{y\in\left(\alpha_j,\beta_j\right)} \left|S_j(y) - S_{j+1}(y)\right| > 0, \qquad j = 1, \ldots, k. \tag{14}$$

We suppose that all measures $\left\{\mathbf{P}_{\boldsymbol{\vartheta}}^{(T)}, \boldsymbol{\vartheta} \in \boldsymbol{\Theta}\right\}$ induced by this process in the space $(\mathcal{C}(0,T), \mathcal{B}(0,T))$ are equivalent to the measure $\mathbf{P}^{(T)}$, which corresponds to the process

$$\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T$$

(see Liptser and Shiryayev 2001). The likelihood ratio

$$L\left(\boldsymbol{\vartheta}, X^T\right) = \frac{\mathrm{d}\mathbf{P}_{\boldsymbol{\vartheta}}^{(T)}}{\mathrm{d}\mathbf{P}^{(T)}}\left(X^T\right), \qquad \boldsymbol{\vartheta} \in \boldsymbol{\Theta},$$

in this problem is the random function

$$\ln L\left(\boldsymbol{\vartheta}, X^T\right) = \sum_{j=1}^{k+1}\int_0^T \frac{S_j(X_t)}{\sigma(X_t)^2}\mathbb{1}_{\{\vartheta_{j-1}<X_t\le\vartheta_j\}}\mathrm{d}X_t - \sum_{j=1}^{k+1}\int_0^T \frac{S_j(X_t)^2}{2\sigma(X_t)^2}\mathbb{1}_{\{\vartheta_{j-1}<X_t\le\vartheta_j\}}\mathrm{d}t.$$

The MLE $\widehat{\boldsymbol{\vartheta}}_T$ is defined by the same equation

$$L\left(\widehat{\boldsymbol{\vartheta}}_T, X^T\right) = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L\left(\boldsymbol{\theta}, X^T\right),$$

where the function $L\left(\boldsymbol{\vartheta}, X^T\right)$ is not differentiable with respect to $\boldsymbol{\vartheta}$.

Note that

$$\mathbb{1}_{\{\vartheta_{j-1}<x\le\vartheta_j\}} = \mathbb{1}_{\{x\le\vartheta_j\}} - \mathbb{1}_{\{x\le\vartheta_{j-1}\}}.$$

Hence

$$\sum_{j=1}^{k+1} S_j(x)\,\mathbb{1}_{\{\vartheta_{j-1}<x\le\vartheta_j\}} = \sum_{j=1}^{k+1} S_j(x)\,\mathbb{1}_{\{x\le\vartheta_j\}} - \sum_{j=1}^{k+1} S_j(x)\,\mathbb{1}_{\{x\le\vartheta_{j-1}\}} = S_{k+1}(x) + \sum_{j=1}^{k}\left[S_j(x) - S_{j+1}(x)\right]\mathbb{1}_{\{x\le\vartheta_j\}}$$

and we can write the likelihood ratio as product of $k+1$ "likelihood ratios"

$$\hat{L}\left(\boldsymbol{\vartheta}, X^T\right) = \frac{\mathrm{d}\mathbf{P}_{\boldsymbol{\vartheta}}^{(T)}}{\mathrm{d}\mathbf{P}_0^{(T)}}\left(X^T\right) = L_{k+1}\left(X^T\right)\prod_{j=1}^{k} L_j\left(\vartheta_j, X^T\right), \tag{15}$$

where

$$\ln L_{k+1}\left(X^T\right)=\int_0^T \frac{S_{k+1}\left(X_t\right)}{\sigma\left(X_t\right)^2}\,\mathrm{d}X_t-\int_0^T \frac{S_{k+1}\left(X_t\right)^2}{2\sigma\left(X_t\right)^2}\,\mathrm{d}t$$

and

$$\begin{aligned}\ln L_j\left(\vartheta_j,X^T\right)=&\int_0^T \frac{S_j\left(X_t\right)-S_{j+1}\left(X_t\right)}{\sigma\left(X_t\right)}\,\mathbb{1}_{\left\{X_t\leq\vartheta_j\right\}}\,\mathrm{d}X_t\\&-\int_0^T \frac{\left[S_j\left(X_t\right)^2-S_{j+1}\left(X_t\right)^2\right]}{2\sigma\left(X_t\right)^2}\mathbb{1}_{\left\{X_t\leq\vartheta_j\right\}}\,\mathrm{d}t.\end{aligned}$$

This allows us to reduce the calculation of the MLE $\widehat{\boldsymbol{\vartheta}}_T$ of multidimensional parameter $\boldsymbol{\vartheta}$ to k one-dimensional problems :

$$\hat\vartheta_{j,T}=\mathrm{argmax}_{\vartheta_j\in\Theta_j}\,L_j\left(\vartheta_j,X^T\right),\qquad j=1,\ldots,k,$$

and to put $\widehat{\boldsymbol{\vartheta}}_T=\left(\hat\vartheta_{1,T},\ldots,\hat\vartheta_{k,T}\right)$.

To introduce the BE $\tilde{\boldsymbol{\vartheta}}_T$ we suppose that $\boldsymbol{\vartheta}$ is a random vector with a known continuous positive density a priori $p\left(\boldsymbol{\theta}\right),\boldsymbol{\theta}\in\boldsymbol{\Theta}$ and the loss function $\ell\left(\boldsymbol{u}\right)\geq 0,\boldsymbol{u}\in\mathscr{R}^k$ is symmetric $\ell\left(-\boldsymbol{u}\right)=\ell\left(\boldsymbol{u}\right)$, strictly convex and has polynomial majorant. The estimator $\widetilde{\boldsymbol{\vartheta}}_T$ is defined as solution of the following equation

$$\int_{\boldsymbol{\Theta}}\mathbf{E}_{\boldsymbol{\theta}}\ell\left(\widetilde{\boldsymbol{\vartheta}}_T-\boldsymbol{\theta}\right)\,p\left(\boldsymbol{\theta}\right)\mathrm{d}\boldsymbol{\theta}=\inf_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\int_{\boldsymbol{\Theta}}\mathbf{E}_{\boldsymbol{\theta}}\ell\left(\boldsymbol{\vartheta}-\boldsymbol{\theta}\right)\,p\left(\boldsymbol{\theta}\right)\mathrm{d}\boldsymbol{\theta}.$$

Remind that in the case $\ell\left(\boldsymbol{u}\right)=\left|\boldsymbol{u}\right|^2$ this estimator is

$$\widetilde{\boldsymbol{\vartheta}}_T=\frac{\int_{\boldsymbol{\Theta}}\boldsymbol{\theta}L\left(\boldsymbol{\theta},X^T\right)p\left(\boldsymbol{\theta}\right)\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}}L\left(\boldsymbol{\theta},X^T\right)p\left(\boldsymbol{\theta}\right)\mathrm{d}\boldsymbol{\theta}}.$$

Let us introduce $\hat{\boldsymbol{u}}_{\boldsymbol{\vartheta}}=\left(\hat u_{1,\boldsymbol{\vartheta}},\ldots,\hat u_{k,\boldsymbol{\vartheta}}\right)$, where

$$\hat u_{j,\boldsymbol{\vartheta}}=\frac{\hat u_j}{\gamma_j\left(\boldsymbol{\vartheta}\right)^2},\quad \gamma_j\left(\boldsymbol{\vartheta}\right)^2=\frac{\left(S_{j+1}\left(\vartheta_j\right)-S_j\left(\vartheta_j\right)\right)^2}{\sigma\left(\vartheta_j\right)^2}\,f\left(\boldsymbol{\vartheta},\vartheta_j\right),\tag{16}$$

and $\hat u_1,\ldots,\hat u_k$ are independent random variables defined by the equalities

$$\hat u_j=\mathrm{argsup}_{u\in\mathscr{R}}\left[W_j\left(u\right)-\frac{1}{2}\left|u\right|\right].$$

Here $W_j\left(\cdot\right),j=1,\ldots,k$ are independent two-sided Wiener processes.

Let us define the random vector $\tilde{\boldsymbol{u}}_{\boldsymbol{\vartheta}}$ as solution of the following equation

$$\int_{\mathscr{R}^k} \ell\left(\tilde{\boldsymbol{u}}_{\boldsymbol{\vartheta}}-\boldsymbol{u}\right) Z(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}=\inf_{\boldsymbol{v}\in\mathscr{R}^k} \int_{\mathscr{R}^k} \ell(\boldsymbol{v}-\boldsymbol{u})\, Z(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u},$$

where

$$Z(\boldsymbol{u})=\exp\left\{\sum_{j=1}^{k}\left[\gamma_j(\boldsymbol{\vartheta})\, W_j\left(u_j\right)-\frac{\left|u_j\right|}{2}\gamma_j(\boldsymbol{\vartheta})^2\right]\right\}. \tag{17}$$

Theorem 1 *Suppose that these conditions $\mathcal{ES}$, $\mathcal{A}$ and (14) are fulfilled, then the MLE $\widehat{\boldsymbol{\vartheta}}_T$ and BE $\widetilde{\boldsymbol{\vartheta}}_T$ are consistent, have the following limit distributions:*

$$T\left(\widehat{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}\right)\Longrightarrow \hat{\boldsymbol{u}}_{\boldsymbol{\vartheta}}, \qquad T\left(\widetilde{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}\right)\Longrightarrow \tilde{\boldsymbol{u}}_{\boldsymbol{\vartheta}}$$

and the moments converge : for any $p>0$

$$\lim_{T\to\infty} T^p\, \mathbf{E}_\vartheta \left|\widehat{\boldsymbol{\vartheta}}_T-\vartheta\right|^p=\mathbf{E}\left|\hat{\boldsymbol{u}}_{\boldsymbol{\vartheta}}\right|^p, \qquad \lim_{T\to\infty} T^p\, \mathbf{E}_\vartheta \left|\tilde{\vartheta}_T-\vartheta\right|^p=\mathbf{E}\left|\tilde{\boldsymbol{u}}_{\boldsymbol{\vartheta}}\right|^p.$$

The proof is given in the Sect. 5.

4 Examples of threshold models

Below we consider several other threshold type ergodic diffusion processes and discuss the properties of parameter estimators for these models.

4.1 Simple threshold model

Suppose that the observed process is

$$\mathrm{d}X_t=\rho_1\, \mathbb{1}_{\{X_t<\vartheta\}}\mathrm{d}t-\rho_2\, \mathbb{1}_{\{X_t\geq\vartheta\}}\mathrm{d}t+\sigma\mathrm{d}W_t, \quad 0\leq t\leq T, \tag{18}$$

where $\rho_i>0$ and $\vartheta\in(\alpha,\beta)$. Then this process is ergodic with exponential type invariant density

$$f(\vartheta,x)=\frac{1}{G(\vartheta)}\exp\left\{-\frac{2\rho(x,\vartheta)\,|x-\vartheta|}{\sigma^2}\right\},$$

where $\rho(x,\vartheta)=\rho_1\mathbb{1}_{\{x<\vartheta\}}+\rho_2\mathbb{1}_{\{x\geq\vartheta\}}$ and $G(\vartheta)$ is the normalizing constant.

The MLE $\hat{\vartheta}_T$ and BE $\tilde{\vartheta}_T$ have the same properties as in Proposition 1 and the corresponding function $\Gamma_\vartheta^2=\sigma^{-4}2\,\rho_2\rho_1(\rho_2+\rho_1)$.

The proof see in the Sect. 5.

4.2 Simple switching

Suppose that in the model (18), we have $\rho_1 = \rho_2 = \rho > 0$. Then, the observed process is

$$\mathrm{d}X_t = -\rho\,\mathrm{sgn}\,(X_t - \vartheta)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad 0 \le t \le T, \tag{19}$$

where $\vartheta \in \Theta = (\alpha, \beta)$. This *Simple Switching Process* was studied in Kutoyants (2004), Sect. 3.4.1. Remind that it has Laplace type invariant density

$$f\,(\vartheta, x) = \frac{\rho}{\sigma^2}\,e^{-\frac{2\rho}{\sigma^2}|x-\vartheta|}.$$

The likelihood ratio formula has the representation

$$L\left(\vartheta, X^T\right) = \exp\left\{-\frac{\rho}{\sigma^2}\int_0^T \mathrm{sgn}\,(X_t - \vartheta)\,\mathrm{d}X_t - \frac{\rho^2 T}{2\sigma^2}\right\}.$$

Hence, the MLE $\hat{\vartheta}_T$ is defined by the equation

$$\int_0^T \mathrm{sgn}\left(X_t - \hat{\vartheta}_T\right)\mathrm{d}X_t = \inf_{\vartheta \in (\alpha,\beta)} \int_0^T \mathrm{sgn}\,(X_t - \vartheta)\,\mathrm{d}X_t.$$

Note that the last stochastic integral we find in Tanaka–Meyer representation of the local time $\Lambda_T\,(x)$ of diffusion process (see Revuz and Yor 1991)

$$\Lambda_T\,(\vartheta) = |X_T - \vartheta| - |X_0 - \vartheta| - \int_0^T \mathrm{sgn}\,(X_t - \vartheta)\,\mathrm{d}X_t$$

and the maximum likelihood is in some sense asymptotically equivalent to the *maximum local time estimator*. Remind that $f_T^\circ\,(x) = \Lambda_T\,(x)\,/\,T\sigma^2$ is the consistent, asymptotically normal and asymptotically efficient (in nonparametric statement) estimator of the invariant density (see Kutoyants 2004 for details), and we have obviously

$$\sup_{\vartheta \in \Theta} f\,(\vartheta_0, \vartheta) = f\,(\vartheta_0, \vartheta_0).$$

We have the same asymptotic properties of the MLE and BE as in the Proposition 1. The proof can be found in Kutoyants (2004), Sect. 3.4.

Note that the *observation window* $(-\infty, \infty)$ can be essentially reduced. Let us put

$$\vartheta^\star_{\sqrt{T}} = \frac{1}{\sqrt{T}}\int_0^{\sqrt{T}} X_t\,\mathrm{d}t.$$

This is an estimator of the method of moments ($\mathbf{E}_\vartheta \xi = \vartheta$). It is consistent and asymptotically normal

$$T^{1/4}\left(\vartheta^\star_{\sqrt{T}} - \theta\right) \Longrightarrow \mathcal{N}\left(0, d^2\left(\vartheta\right)\right)$$

(see Kutoyants (2004), p. 270, where $d^2\left(\vartheta\right)$ is calculated). Introduce the *window*

$$\mathbb{B}_T = \left[\vartheta^\star_{\sqrt{T}} - T^{-1/8}, \vartheta^\star_{\sqrt{T}} + T^{-1/8}\right].$$

The MLE and BE we define with the help of the following LR

$$L\left(\vartheta, X^T_{\sqrt{T}}\right) = \exp\left\{-\frac{\rho}{\sigma^2}\int_{\sqrt{T}}^T \operatorname{sgn}(X_t - \vartheta)\mathbb{1}_{\{X_t\in\mathbb{B}_T\}}\mathrm{d}X_t \right.$$
$$\left. -\frac{\rho^2}{2\sigma^2}\int_{\sqrt{T}}^T \mathbb{1}_{\{X_t\in\mathbb{B}_T\}}\mathrm{d}t\right\}.$$

Then, these estimators have the same asymptotic properties as if the observation window is $\mathbb{B}_T = (-\infty, \infty)$.

This a bit surprising result is probably typical for singular estimation problems. The analyse of the proof of the properties of estimators (see Kutoyants 2004, Sect. 3.4) shows that only the values of X_t close to the true value ϑ_0 have contribution to the limit likelihood ratio. Hence, all other observations are irrelevant and can be deleted by introducing this window. This is a general property of threshold models and the similar construction can be done in the case of all other threshold models studied in this work.

4.3 Multi threshold O–U process

Suppose that the observed process is

$$\mathrm{d}X_t = -\sum_{l=1}^{k+1}\rho_l\, X_t\, \mathbb{1}_{\{\vartheta_{l-1}<X_t\leq\vartheta_l\}}\mathrm{d}t + \sigma\mathrm{d}W_t, \quad 0\leq t\leq T, \tag{20}$$

where $\rho_1 > 0$, $\rho_{k+1} > 0$, $\rho_l \neq \rho_m > 0$, $\vartheta_0 = -\infty$, $\vartheta_{k+1} = \infty$ and $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_k) \in \boldsymbol{\Theta} = \Theta_1 \times \cdots \times \Theta_k$, $\Theta_l = (\alpha_l, \beta_l)$, $\beta_l < \alpha_{l+1}$. Then this process is a particular case of (12) and conditions of ergodicity and identifiability are fulfilled. Therefore we obtain the mentioned in the Theorem 1 properties of the estimators for the details, see Sect. 5.

5 Proofs

First note that the parameter estimation problems for the models of the observations (3), (18)–(20) are particular cases of the threshold estimation problem for stochastic process (12). Therefore, it is sufficient to prove the Theorem 1.

5.1 Proof of Theorem 1

The proof of this theorem is based on the two remarkable theorems by (Ibragimov and Khasminskii 1981; Theorems 1.10.1, 1.10.2) and some results obtained before in Kutoyants (2004). Let us remind the main steps of this approach. Introduce the random function (normalized likelihood ratio)

$$Z_T\left(\boldsymbol{u}\right)=\frac{L\left(\boldsymbol{\vartheta}+\frac{\boldsymbol{u}}{T},X^T\right)}{L\left(\boldsymbol{\vartheta},X^T\right)},\qquad \boldsymbol{u}\in\boldsymbol{U}_T=U_{1,T}\times\cdots\times U_{k,T},$$

where $U_{j,T}=\left(T\left(\alpha_j-\vartheta_j\right),T\left(\beta_j-\vartheta_j\right)\right)$. The properties of estimators follow, roughly speaking, from the weak convergence of this function to the limit random field (17): $Z_T\left(\boldsymbol{u}\right)\Longrightarrow Z\left(\boldsymbol{u}\right)$.

Suppose that we have already this convergence and (for simplicity) assume that $k=1$ (for the multidimensional case see. Then for the MLE we have (ϑ is the true value):

$$\begin{aligned}
&\mathbf{P}_\vartheta\left\{T\left(\hat\vartheta_T-\vartheta\right)<x\right\}\\
&=\mathbf{P}\left\{\sup_{T(\theta-\vartheta)<x}L\left(\theta,X^T\right)>\sup_{T(\theta-\vartheta)\geq x}L\left(\theta,X^T\right)\right\}\\
&=\mathbf{P}\left\{\sup_{T(\theta-\vartheta)<x}\frac{L\left(\vartheta,X^T\right)}{L\left(\vartheta_0,X^T\right)}>\sup_{T(\theta-\vartheta)\geq x}\frac{L\left(\vartheta,X^T\right)}{L\left(\vartheta_0,X^T\right)}\right\}\\
&=\mathbf{P}\left\{\sup_{u<x}Z_T\left(u\right)>\sup_{u\geq x}Z_T\left(u\right)\right\}\\
&\longrightarrow\mathbf{P}\left\{\sup_{u<x}Z\left(u\right)>\sup_{u\geq x}Z\left(u\right)\right\}\\
&=\mathbf{P}\left(\frac{\hat u}{\gamma\left(\vartheta\right)^2}<x\right),\quad\text{i.e.}\quad T\left(\hat\vartheta_T-\vartheta\right)\Longrightarrow\frac{\hat u}{\gamma\left(\vartheta\right)^2}.
\end{aligned}\tag{21}$$

where we put $\theta=\vartheta+T^{-1}u$ and $\gamma_1\left(\vartheta\right)=\gamma\left(\vartheta\right)$.

To describe the behavior of the BE we take we for simplicity the square loss function and use the same change of variables $\theta=\vartheta+u/T\equiv\theta_u$ and,

$$\tilde{\vartheta}_T = \frac{\int_\alpha^\beta \theta p(\theta) L\left(\theta, X^T\right) \mathrm{d}\theta}{\int_\alpha^\beta p(\theta) L\left(\theta, X^T\right) \mathrm{d}\theta} = \vartheta + \frac{1}{T} \frac{\int_{U_T} up(\theta_u) L\left(\theta_u, X^T\right) \mathrm{d}u}{\int_{U_T} p(\theta_u) L\left(\theta_u, X^T\right) \mathrm{d}u}$$

$$= \vartheta + \frac{1}{T} \frac{\int_{U_T} up(\theta_u) \frac{L(\theta_u, X^T)}{L(\vartheta, X^T)} \mathrm{d}u}{\int_{U_T} p(\theta_u) \frac{L(\theta_u, X^T)}{L(\vartheta, X^T)} \mathrm{d}u} = \vartheta + \frac{1}{T} \frac{\int_{U_T} up(\theta_u) Z_T(u) \, \mathrm{d}u}{\int_{U_T} p(\theta_u) Z_T(u) \, \mathrm{d}u}.$$

Then, using the convergence $p(\theta_u) \to p(\vartheta)$, we can write

$$\mathbf{P}_\vartheta \left\{ T\left(\tilde{\vartheta}_T - \vartheta\right) < x \right\} = \mathbf{P}\left\{ \frac{\int_{U_T} u\; p(\theta_u) Z_T(u) \, \mathrm{d}u}{\int_{U_T} p(\theta_u) Z_T(u) \, \mathrm{d}u} < x \right\}$$
$$\longrightarrow \mathbf{P}\left\{ \frac{\int_R u\; Z(u) \, \mathrm{d}u}{\int_R Z(u) \, \mathrm{d}u} < x \right\} = \mathbf{P}\left(\frac{\tilde{u}}{\gamma(\vartheta)^2} < x \right). \qquad (22)$$

The random variables $\hat{u}$ and $\tilde{u}$ are defined in (5).

We see that to prove the theorem we need to prove the convergences (21), (22). These convergences together with the estimates on the large deviations of estimators will provide the convergence of moments. The corresponding sufficient conditions are given in the mentioned above theorems by Ibragimov and Khasminskii. Let us introduce the conditions

A. The finite dimensional distributions of the random function $Z_T(\cdot)$ converge to the finite dimensional distributions of the function $Z(\cdot)$.
B. There exist constants $B > 0, m > 0, b > 0$ and d such that for any $R > 0$ and $|\boldsymbol{u}| \le R, |\boldsymbol{v}| \le R$

$$\mathbf{E}_{\boldsymbol{\vartheta}} \left| Z_T^{\frac{1}{2m}}(\boldsymbol{u}) - Z_T^{\frac{1}{2m}}(\boldsymbol{v}) \right|^{2m} \le B\left(1 + R^b\right) |\boldsymbol{u} - \boldsymbol{v}|^d. \qquad (23)$$

C. For any $N > 0$, there exists constant $C_N > 0$, such that

$$\mathbf{E}_{\boldsymbol{\vartheta}} Z_T^{\frac{1}{2}}(\boldsymbol{u}) \le \frac{C_N}{|\boldsymbol{u}|^N}. \qquad (24)$$

These conditions are the version of the conditions of Theorems 1.10.1 (with $d > k$) and 1.10.2 Ibragimov and Khasminskii (1981), which we will verify in this work.

We start with the condition **A**. Let us consider the case when all $u_j > 0$ and denote $h_j(x) = S_j(x)/\sigma(x)$. Note that

$$\mathbb{1}_{\left\{\vartheta_{j-1} + \frac{u_{j-1}}{T} < X_t \le \vartheta_j + \frac{u_j}{T}\right\}} - \mathbb{1}_{\{\vartheta_{j-1} < X_t \le \vartheta_j\}}$$
$$= \mathbb{1}_{\left\{\vartheta_j < X_t \le \vartheta_j + \frac{u_j}{T}\right\}} - \mathbb{1}_{\left\{\vartheta_{j-1} < X_t \le \vartheta_{j-1} + \frac{u_{j-1}}{T}\right\}} = \mathbb{1}_{\{\mathbb{B}_j\}} - \mathbb{1}_{\{\mathbb{B}_{j-1}\}}$$

in obvious notation.

Then, the likelihood ratio $Z_T(\boldsymbol{u})$ can be written as follows

$$\ln Z_T(\boldsymbol{u}) = \sum_{j=1}^{k+1} \int_0^T h_j(X_t)\left[\mathbb{1}_{\{\mathbb{B}_j\}} - \mathbb{1}_{\{\mathbb{B}_{j-1}\}}\right] \mathrm{d}W_t$$
$$-\frac{1}{2}\sum_{j=1}^{k} \int_0^T \left[h_j(X_t) - h_{j+1}(X_t)\right]^2 \mathbb{1}_{\{\mathbb{B}_j\}} \mathrm{d}t.$$

Using the local time estimator $f_T^\circ(x)$ of the invariant density $f(\boldsymbol{\vartheta}, x)$ we write

$$\begin{aligned}
&\int_0^T \left[h_j(X_t) - h_{j+1}(X_t)\right]^2 \mathbb{1}_{\{\mathbb{B}_j\}} \mathrm{d}t \\
&= T \int_{-\infty}^{\infty} \left[h_j(x) - h_{j+1}(x)\right]^2 \mathbb{1}_{\left\{\vartheta_j < x \leq \vartheta_j + \frac{u_j}{T}\right\}} f_T^\circ(x)\, \mathrm{d}x \\
&= T \int_{\vartheta_j}^{\vartheta_j + \frac{u_j}{T}} \left[h_j(x) - h_{j+1}(x)\right]^2 f_T^\circ(x)\, \mathrm{d}x \\
&= T \int_{\vartheta_j}^{\vartheta_j + \frac{u_j}{T}} \left[h_j(x) - h_{j+1}(x)\right]^2 f(\boldsymbol{\vartheta}, x)\, \mathrm{d}x \\
&\quad + T \int_{\vartheta_j}^{\vartheta_j + \frac{u_j}{T}} \left[h_j(x) - h_{j+1}(x)\right]^2 \left[f_T^\circ(x) - f(\boldsymbol{\vartheta}, x)\right] \mathrm{d}x.
\end{aligned}$$

For the random function $\eta_T(x) = T\left(f_T^\circ(x) - f(\boldsymbol{\vartheta}, x)\right)$ we have the estimate: for any $p > 0$ there exist constants $C_* > 0$ and $c_* > 0$ such that

$$\mathbf{E}_{\boldsymbol{\vartheta}} \left|\eta_T(x)\right|^p \leq C_*\, e^{-c_*|x|} \tag{25}$$

see Proposition 1.11 in Kutoyants (2004). This estimate allows us to prove that the last integral tends to zero as $T \to \infty$. We have as well

$$\begin{aligned}
&T \int_{\vartheta_j}^{\vartheta_j + \frac{u_j}{T}} \left[h_j(x) - h_{j+1}(x)\right]^2 f(\boldsymbol{\vartheta}, x)\, \mathrm{d}x \\
&\quad \longrightarrow u_j \left[h_j(\vartheta_j) - h_{j+1}(\vartheta_j)\right]^2 f(\boldsymbol{\vartheta}, \vartheta_j) = u_j \gamma_j(\boldsymbol{\vartheta})^2.
\end{aligned}$$

Therefore,

$$\sum_{j=1}^{k} \int_0^T \left[h_j(X_t) - h_{j+1}(X_t)\right]^2 \mathbb{1}_{\{\mathbb{B}_j\}} \mathrm{d}t \longrightarrow \sum_{j=1}^{k} u_j \gamma_j(\boldsymbol{\vartheta})^2.$$

The central limit theorem for stochastic integrals yields the asymptotic normality of the vector $\boldsymbol{\xi}_T = \left(\xi_{1,T}, \ldots, \xi_{k,T}\right)$

$$\xi_{j,T} = \int_0^T \left[h_j\left(X_t\right) - h_{j+1}\left(X_t\right)\right] \mathbb{1}_{\{\mathbb{B}_j\}} \mathrm{d}W_t \Longrightarrow \mathcal{N}\left(0, u_j \gamma_j\left(\boldsymbol{\vartheta}\right)^2\right)$$

with asymptotically independent components, because

$$\mathbf{E}_{\boldsymbol{\vartheta}} \xi_{j,T} \xi_{l,T} = 0, \qquad l \neq j.$$

Moreover, if we put $\xi_{j,T} = \xi_{j,T}\left(u_j\right)$ and consider the vector $\boldsymbol{\xi}_{j,T} = \left(\xi_{j,T}\left(u_{j,1}\right), \ldots, \xi_{j,T}\left(u_{j,n}\right)\right)$, where $u_{j,1}, \ldots, u_{j,n}$ is some collection of values from $U_{j,T}$, then

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\vartheta}} \xi_{j,T}\left(u_{j,r}\right) \xi_{j,T}\left(u_{j,q}\right) &= T \int_{\vartheta_j}^{\vartheta_j + \frac{u_{j,r} \wedge u_{j,q}}{T}} \left[h_j(x) - h_{j+1}(x)\right]^2 f\left(\boldsymbol{\vartheta}, x\right) \mathrm{d}x \\ &\longrightarrow \left[u_{j,r} \wedge u_{j,q}\right] \gamma_j\left(\boldsymbol{\vartheta}\right)^2.\end{aligned}$$

Using this equality and preceding limits, we can show the convergence

$$\left(\xi_{j,T}\left(u_{j,1}\right), \ldots, \xi_{j,T}\left(u_{j,n}\right)\right) \Longrightarrow \gamma_j\left(\boldsymbol{\vartheta}\right)\left(W_j\left(u_{j,1}\right), \ldots, W_j\left(u_{j,1}\right)\right).$$

Therefore, the condition **A** is fulfilled.

To verify **B**, we do it twice. The first time we check this condition with $m = 1$, which is sufficient for BEs (multidimensional case) and then (for MLE), we verify it for the partial likelihoods $Z_{j,T}(u)$. Following Kutoyants (2004), Lemma 3.28, we write (we suppose that $v_j < u_j$)

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\vartheta}} \left|Z_T^{1/2}(\boldsymbol{u}) - Z_T^{1/2}(\boldsymbol{v})\right|^2 &\leq \frac{1}{4} \sum_{j=1}^k \mathbf{E}_* \int_0^T \left[h_j\left(X_t\right) - h_{j+1}\left(X_t\right)\right]^2 \mathbb{1}_{\left\{\tilde{\mathbb{B}}_j\right\}} \mathrm{d}t \\ &= \frac{1}{4} \sum_{j=1}^k T \int_{-\infty}^{\infty} \left[h_j(x) - h_{j+1}(x)\right]^2 \mathbb{1}_{\left\{\vartheta_j + \frac{v_j}{T} < x \leq \vartheta_j + \frac{u_j}{T}\right\}} f_*(x) \mathrm{d}x \\ &= \frac{1}{4} \sum_{j=1}^k T \int_{\vartheta_j + \frac{vj}{T}}^{\vartheta_j + \frac{u_j}{T}} \left[h_j(x) - h_{j+1}(x)\right]^2 f_*(x) \mathrm{d}x \\ &\leq C \sum_{j=1}^k \left|u_j - v_j\right| \leq C \left\|\boldsymbol{u} - \boldsymbol{v}\right\|. \qquad (26)\end{aligned}$$

Here $\mathbf{E}_*$ and $f_*(\cdot)$ are expectation and invariant density which correspond to the stochastic differential equation

$$\begin{aligned}\mathrm{d}X_t =& \sum_{j=1}^{k+1} S_j\left(X_t\right)\left[\mathbb{1}_{\left\{\vartheta_{j-1}+\frac{u_{j-1}}{T}<X_t\leq\vartheta_j+\frac{u_j}{T}\right\}}+\mathbb{1}_{\left\{\vartheta_{j-1}+\frac{v_{j-1}}{T}<X_t\leq\vartheta_j+\frac{v_j}{T}\right\}}\right]\mathrm{d}t\\ &+\sigma\left(X_t\right)\mathrm{d}W_t, \qquad X_0, \quad 0\leq t\leq T\end{aligned}$$

(see details in Kutoyants 2004, p. 379).

The condition **B** in the case of the study the MLE we check for the components $Z_{j,T}\left(u_j\right), u_j\in\mathbf{U}_{j,T}$ separately as follows. Let us introduce the stochastic process

$$V_{j,t}=\left(\frac{Z_{j,t}\left(u_j\right)}{Z_{j,t}\left(v_j\right)}\right)^{1/16}, \quad V_{j,0}=1, \quad 0\leq t\leq T$$

and denote

$$g_j\left(x\right)=\frac{S_j\left(x\right)-S_{j+1}\left(x\right)}{\sigma\left(x\right)}.$$

Then the process

$$\begin{aligned}V_{j,t}=\exp\Bigg\{&\frac{1}{16}\int_0^t\frac{S_j\left(X_s\right)-S_{j+1}\left(X_s\right)}{\sigma\left(X_s\right)^2}\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\mathrm{d}X_s\\ &-\frac{1}{32}\int_0^t\frac{S_j\left(X_s\right)^2-S_{j+1}\left(X_s\right)^2}{\sigma\left(X_s\right)^2}\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\mathrm{d}s\Bigg\}\end{aligned}$$

by Itô formula admits the representation (under measure $\mathbf{P}_{\boldsymbol{\vartheta}}^{(T)}$)

$$\begin{aligned}V_{j,T}=1+&\frac{1}{16}\int_0^T V_{j,t}\,\frac{S_j\left(X_s\right)-S_{j+1}\left(X_s\right)}{\sigma\left(X_s\right)}\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\,\mathrm{d}W_t\\ &-\frac{15}{512}\int_0^T V_{j,t}\left(\frac{S_j\left(X_s\right)-S_{j+1}\left(X_s\right)}{\sigma\left(X_s\right)}\right)^2\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\,\mathrm{d}t,\end{aligned}$$

Remind that

$$\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<x\leq\vartheta_j+\frac{u_j}{T}\right\}}\sum_{l=1}^{k+1}S_l\left(x\right)\mathbb{1}_{\left\{\vartheta_{l-1}<x\leq\vartheta_l\right\}}=S_{j+1}\left(x\right)\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<x\leq\vartheta_j+\frac{u_j}{T}\right\}}.$$

Therefore we can write

$$\begin{aligned}\mathbf{E}_{\vartheta}\left|Z_{j,T}^{1/16}\left(u_j\right)-Z_{j,T}^{1/16}\left(v_j\right)\right|^4&=\mathbf{E}_{\vartheta}Z_{j,T}^{1/4}\left(v_j\right)\left|1-V_{j,T}\right|^4\\&\leq\left(\mathbf{E}_{\vartheta}Z_{j,T}^{1/2}\left(v_j\right)\right)^{1/2}\left(\mathbf{E}_{\vartheta}\left|1-V_{j,T}\right|^8\right)^{1/2}\\&\leq\left(\mathbf{E}_{\vartheta}\left|1-V_{j,T}\right|^8\right)^{1/2}\end{aligned}$$

because $\mathbf{E}_{\vartheta}Z_{j,T}^{1/2}\left(v_j\right)\leq 1$. Further

$$\begin{aligned}\mathbf{E}_{\vartheta}\left|1-V_{j,T}\right|^8&\leq C_1\,\mathbf{E}_{\vartheta}\left(\int_0^T V_{j,t}\;g_j\left(X_t\right)^2\;\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\;\mathrm{d}t\right)^8\\&\quad+C_2\,\mathbf{E}_{\vartheta}\left(\int_0^T V_{j,t}\;g_j\left(X_t\right)\;\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\;\mathrm{d}W_t\right)^8.\end{aligned}\tag{27}$$

For the last (stochastic) integral we have the estimates

$$\begin{aligned}&\mathbf{E}_{\vartheta}\left(\int_0^T V_{j,t}\;g_j\left(X_t\right)\;\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\;\mathrm{d}W_t\right)^8\\&\quad\leq C\mathbf{E}_{\vartheta}\left(\int_0^T V_{j,t}^2\;g_j\left(X_t\right)^2\;\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\;\mathrm{d}t\right)^4\\&\quad\leq C\mathbf{E}_{\vartheta}\sup_{0\leq t\leq T}V_{j,t}^8\left(\int_0^T g_j\left(X_t\right)^2\;\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\;\mathrm{d}t\right)^4\\&\quad\leq C\left(\mathbf{E}_{\vartheta}\sup_{0\leq t\leq T}V_{j,t}^{16}\right)^{1/2}\\&\qquad\times\left(\mathbf{E}_{\vartheta}\left(\int_0^T g_j\left(X_t\right)^2\;\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\;\mathrm{d}t\right)^8\right)^{1/2}.\end{aligned}$$

Remind that V_t^{16} is martingale and $\mathbf{E}_{\vartheta}V_T^{16}=1$. Using once more the local time estimator of the density we write

$$\int_0^T g_j\left(X_t\right)^2\;\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\mathrm{d}t=T\int_{\vartheta_j+\frac{v_j}{T}}^{\vartheta_j+\frac{u_j}{T}}g_j\left(x\right)^2\;f_T^{\circ}\left(x\right)\mathrm{d}x.$$

Hence

$$\mathbf{E}_{\boldsymbol{\vartheta}}\left(\int_0^T g_j\left(X_t\right)^2 \mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\mathrm{d}t\right)^8$$

$$\leq\left(u_j-v_j\right)^7 T\int_{\vartheta_j+\frac{v_j}{T}}^{\vartheta_j+\frac{u_j}{T}} g_j\left(x\right)^{16}\,\mathbf{E}_{\boldsymbol{\vartheta}}\, f_T^{\circ}\left(x\right)^8\mathrm{d}x\leq C\left(u_j-v_j\right)^8.$$

The expectation $\mathbf{E}_{\boldsymbol{\vartheta}}\, f_T^{\circ}\left(x\right)^8$ due to the estimate (25) is a bounded function. For the first integral in (27), the similar calculations yield the estimate

$$\mathbf{E}_{\boldsymbol{\vartheta}}\left(\int_0^T V_{j,t}\; g_j\left(X_t\right)^2\;\mathbb{1}_{\left\{\vartheta_j+\frac{v_j}{T}<X_s\leq\vartheta_j+\frac{u_j}{T}\right\}}\;\mathrm{d}t\right)^8\leq C\left(u_j-v_j\right)^8.$$

Therefore, for $\left|u_j\right|\leq R$, $\left|v_j\right|\leq R$

$$\mathbf{E}_{\boldsymbol{\vartheta}}\left|Z_{j,T}^{1/16}\left(u_j\right)-Z_{j,T}^{1/16}\left(v_j\right)\right|^8\leq C\left(u_j-v_j\right)^2+\left(u_j-v_j\right)^4$$

$$\leq C\left(1+R^2\right)\left|u_j-v_j\right|^2. \tag{28}$$

To verify condition **C**, we follow the proof of the Lemmas 3.29 and 2.11 in Kutoyants (2004). By condition (14), we have

$$\mathbf{E}_{\boldsymbol{\vartheta}}\sum_{j=1}^k\int_0^T\left[h_j\left(X_t\right)-h_{j+1}\left(X_t\right)\right]^2\mathbb{1}_{\left\{\vartheta_j<X_t\leq\vartheta_j+\delta_j\right\}}\mathrm{d}t$$

$$=T\sum_{j=1}^k\int_{\vartheta_j}^{\vartheta_j+\delta_j}\left[h_j\left(x\right)-h_{j+1}\left(x\right)\right]^2 f\left(\boldsymbol{\vartheta},x\right)\;\mathrm{d}x$$

$$=T\sum_{j=1}^k\kappa_j\delta_j\left(1+o\left(1\right)\right)\geq\kappa T\left|\boldsymbol{\delta}\right|$$

with some positive constants κ,κ_j. Here $\boldsymbol{\delta}=\left(\delta_1,\ldots,\delta_k\right)$ and we suppose for simplicity that all $\delta_j>0$. Hence, the inequality (24) follows from the mentioned above lemmas.

The properties of BE follow from the Theorem 1.10.2 in Ibragimov and Khasminskii (1981) because the conditions **A**, (26) and (24) are sufficient for this theorem.

For the MLE, we do not apply directly the Theorem 1.10.1 in Ibragimov and Khasminskii (1981) because it requires in condition **B** that $d>k$. We follow the modification of this theorem discussed in the proof of the Proposition 2.40 in Kutoyants (2004). Let us consider the vector of likelihood ratios $\boldsymbol{Y}\left(\boldsymbol{u}\right)_T=\left(Z_{1,T}^{1/4}(u_1),\ldots,Z_{k,T}^{1/4}(u_1)\right)$. For the components $Z_{j,T}^{1/4}\left(u_j\right)$, $j=1,\ldots,k$ we have the joint convergence of its dimensional distributions to the distribution of the limit random field

$\boldsymbol{Y}(\boldsymbol{u}) = (Z_1^{1/4}(u_1), \ldots, Z_k^{1/4}(u_1))$ with independent components and the conditions **B** and **C**. Therefore, we have the tightness of the corresponding vector of measures and for each component we have the large deviations estimates: for any $L > 0$ and $N > 0$ there exists $C_N > 0$ such that

$$\mathbf{P}_{\boldsymbol{\vartheta}}^{(T)} \left\{ \sup_{|u_j|>L} Z_{j,T}^{1/4}\left(u_j\right) \geq \frac{1}{L^N} \right\} \leq \frac{C_N}{L^N}.$$

These estimates and the factorization of the likelihood ratio (15) allows us to finish the proof of the properties of MLE mentioned in Theorem 1. Note that the MLE $\hat{\vartheta}_{j,T}$ can be written as

$$\hat{\vartheta}_{j,T} = \operatorname{argmax}_{\theta_j \in \Theta_j} L_j^{1/4}\left(\theta_j, X^T\right)$$

too. □

5.2 Proof of Proposition 2

To prove the Proposition 2, we consider the normalized likelihood ratio (we take $u > 0$)

$$\begin{aligned}
\ln Z_T\left(v, w, u\right) &= \ln \frac{L\left(\vartheta_1 + \frac{v}{\sqrt{T}}, \vartheta_2 + \frac{w}{\sqrt{T}}, \vartheta_3 + \frac{u}{T}, X^T\right)}{L\left(\vartheta_1, \vartheta_2, \vartheta_3, X^T\right)} \\
&= -\frac{v}{\sigma\sqrt{T}} \int_0^T X_t \mathbb{1}_{\{X_t<\vartheta_3\}} \mathrm{d}W_t - \frac{w}{\sigma\sqrt{T}} \int_0^T X_t \mathbb{1}_{\{X_t\geq\vartheta_3\}} \mathrm{d}W_t \\
&\quad + \left(\vartheta_2 - \vartheta_1 + \frac{w-v}{\sqrt{T}}\right) \frac{1}{\sigma} \int_0^T X_t \mathbb{1}_{\left\{\vartheta_3<X_t\leq\vartheta_3+\frac{u}{\sqrt{T}}\right\}} \mathrm{d}W_t \\
&\quad - \frac{1}{\sigma^2} \int_0^T \left[-\frac{v}{\sqrt{T}} \mathbb{1}_{\{X_t<\vartheta_3\}} - \frac{w}{\sqrt{T}} \mathbb{1}_{\{X_t\geq\vartheta_3\}} \right. \\
&\quad \left. + \left(\vartheta_2 - \vartheta_1 + \frac{w-v}{\sqrt{T}}\right) 1_{\left\{\vartheta_3<X_t\leq\vartheta_3+\frac{u}{\sqrt{T}}\right\}} \right]^2 X_t^2 \mathrm{d}t \\
&\equiv v\Delta_{1,T} + w\Delta_{2,T} + \left(\frac{\vartheta_2-\vartheta_1}{\sigma} + \frac{w-v}{\sigma\sqrt{T}}\right) \Delta_{3,T}\left(u\right) - \frac{1}{2}\, J_T,
\end{aligned}$$

where the last equality introduce the notation for these integrals. For the last integral we can write

$$\begin{aligned}
J_T &= \frac{v^2}{\sigma^2 T} \int_0^T X_t^2 \mathbb{1}_{\{X_t<\vartheta_3\}} \mathrm{d}t + \frac{w^2}{\sigma^2 T} \int_0^T X_t^2 \mathbb{1}_{\{X_t\geq\vartheta_3\}} \mathrm{d}t \\
&\quad + \frac{(\vartheta_2-\vartheta_1)^2}{\sigma^2} \int_0^T X_t^2 \mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u}{T}\}} \mathrm{d}t + o\left(1\right).
\end{aligned} \tag{29}$$

For the first two integrals by the law of large numbers we have

$$\frac{1}{T}\int_0^T X_t^2\,\mathbb{1}_{\{X_t<\vartheta_3\}}\,\mathrm{d}t \longrightarrow \mathbf{E}_{\boldsymbol{\vartheta}}\xi^2\mathbb{1}_{\{\xi<\vartheta_3\}}, \tag{30}$$

$$\frac{1}{T}\int_0^T X_t^2\,\mathbb{1}_{\{X_t\geq\vartheta_3\}}\,\mathrm{d}t \longrightarrow \mathbf{E}_{\boldsymbol{\vartheta}}\xi^2\mathbb{1}_{\{\xi\geq\vartheta_3\}}, \tag{31}$$

and for the last one using the local time estimator of the density we obtain

$$\int_0^T X_t^2\mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u}{T}\}}\mathrm{d}t = T\int_{\vartheta_3}^{\vartheta_3+\frac{u}{T}} x^2\, f_T^\circ(x)\,\mathrm{d}x = T\int_{\vartheta_3}^{\vartheta_3+\frac{u}{T}} x^2\, f(\boldsymbol{\vartheta},x)\,\mathrm{d}x$$
$$+T\int_{\vartheta_3}^{\vartheta_3+\frac{u}{T}} x^2\,\left(f_T^\circ(x)-f(\boldsymbol{\vartheta},x)\right)\,\mathrm{d}x = u\,\vartheta_3^2\,f(\boldsymbol{\vartheta},\vartheta_3)+o(1),$$

where in $o(1)$ we used once more the estimate (25). Therefore,

$$J_T \longrightarrow \frac{v^2}{\sigma^2}\,\mathbf{E}_{\boldsymbol{\vartheta}}\xi^2\mathbb{1}_{\{\xi\leq\vartheta_3\}} + \frac{w^2}{\sigma^2}\,\mathbf{E}_{\boldsymbol{\vartheta}}\xi^2\mathbb{1}_{\{\xi\geq\vartheta_3\}} + u\,\frac{(\vartheta_2-\vartheta_1)^2\,\vartheta_3^2}{\sigma^2}\,f(\boldsymbol{\vartheta},\vartheta_3)\,.$$

For the stochastic integrals $\Delta_{1,T}$ and $\Delta_{2,T}$ from (30), (31) and by the central limit theorem we have the convergence

$$\Delta_{1,T} \Longrightarrow \zeta_1 \sim \mathcal{N}(0,I_1),\quad I_1=\frac{1}{\sigma^2}\,\mathbf{E}_{\boldsymbol{\vartheta}}\xi^2\mathbb{1}_{\{\xi\leq\vartheta_3\}} \tag{32}$$

$$\Delta_{2,T} \Longrightarrow \zeta_2 \sim \mathcal{N}(0,I_2),\quad I_2=\frac{1}{\sigma^2}\,\mathbf{E}_{\boldsymbol{\vartheta}}\xi^2\mathbb{1}_{\{\xi\geq\vartheta_3\}}, \tag{33}$$

where the random variables ζ_1 and ζ_2 are independent.

Let us consider $\Delta_T=\lambda_1\Delta_{3,T}(u_1)+\lambda_2\Delta_{3,T}(u_2)$. We have

$$\Delta_T=\int_0^T\left[\lambda_1X_t\mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u_1}{T}\}}+\lambda_2X_t\mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u_2}{T}\}}\right]\mathrm{d}W_t.$$

Note that

$$\int_0^T\left[\lambda_1X_t\mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u_1}{T}\}}+\lambda_2X_t\mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u_2}{T}\}}\right]^2\mathrm{d}t$$
$$=\lambda_1^2\int_0^T X_t^2\mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u_1}{T}\}}\mathrm{d}t+\lambda_2^2\int_0^T X_t^2\mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u_2}{T}\}}\mathrm{d}t$$
$$+2\lambda_1\lambda_2\int_0^T X_t^2\mathbb{1}_{\left\{\vartheta_3<X_t\leq\vartheta_3+\frac{u_1\wedge u_2}{T}\right\}}\mathrm{d}t$$
$$\longrightarrow\left[u_1\,\lambda_1^2+u_2\,\lambda_2^2+2\lambda_1\lambda_2\,(u_1\wedge u_2)\right]\vartheta_3^2\;f(\boldsymbol{\vartheta},\vartheta_3)\equiv d^2.$$

Hence Δ_T is asymptotically normal $\Delta_T \Rightarrow \Delta$ with the limit variance d^2. Remind that the same variance has the random variable

$$\Delta = \lambda_1 \vartheta_3 \sqrt{f\left(\boldsymbol{\vartheta}, \vartheta_3\right)}\, W\left(u_1\right) + \lambda_2 \vartheta_3 \sqrt{f\left(\boldsymbol{\vartheta}, \vartheta_3\right)}\, W\left(u_2\right),$$

where $W\left(\cdot\right)$ is a Wiener process. Therefore we have the convergence of the finite dimensional distributions of $\Delta_{3,T}\left(u\right)$ to the finite dimensional distributions of the process $\vartheta_3 \sqrt{f\left(\boldsymbol{\vartheta}, \vartheta_3\right)}\, W\left(u\right)$:

$$\begin{aligned} &\left(\Delta_{3,T}\left(u_1\right), \ldots, \Delta_{3,T}\left(u_k\right)\right) \\ &\quad \Longrightarrow \left(\vartheta_3 \sqrt{f\left(\boldsymbol{\vartheta}, \vartheta_3\right)}\, W\left(u_1\right), \ldots, \vartheta_3 \sqrt{f\left(\boldsymbol{\vartheta}, \vartheta_3\right)}\, W\left(u_k\right)\right). \end{aligned} \tag{34}$$

This convergence together with (32) and (33) allows to write the likelihood ratio random field as

$$\begin{aligned} Z_T\left(v, w, u\right) = \exp\Bigg\{ & v\Delta_{1,T} - \frac{v^2}{2}\mathrm{I}_1 + w\Delta_{2,T} - \frac{w^2}{2}\mathrm{I}_2 \\ & + \left(\frac{\vartheta_2 - \vartheta_1}{\sigma}\right)\Delta_{3,T}\left(u\right) - \frac{|u|}{2}\gamma\left(\boldsymbol{\vartheta}\right)^2 + o\left(1\right)\Bigg\}. \end{aligned}$$

where $\Delta_{1,T}$ and $\Delta_{2,T}$ are asymptotically normal, and

$$\gamma\left(\boldsymbol{\vartheta}\right)^2 = \frac{\left(\vartheta_2 - \vartheta_1\right)^2 \vartheta_3^2}{\sigma^2}\, f\left(\boldsymbol{\vartheta}, \vartheta_3\right) \equiv \gamma^2.$$

Therefore we have the convergence of the finite dimensional distributions of Z_T $\left(v, w, u\right)$ to that of the random function

$$Z\left(v, w, u\right) = e^{v\zeta_1 - \frac{v^2}{2}\mathrm{I}_1}\; e^{w\zeta_2 - \frac{w^2}{2}\mathrm{I}_2}\; e^{\gamma W(u) - \frac{|u|}{2}\gamma^2}, \quad v, w, u \in \mathscr{R}^3.$$

where ζ_1, ζ_2 and $W\left(\cdot\right)$ are independent.

To check the condition **B** in the case of Bayesian estimation we following (26) write $\left(u_2 > u_1 > 0\right)$

$$\begin{aligned} &\mathbf{E}_{\boldsymbol{\vartheta}} \left| Z_T^{1/2}\left(v_1, w_1, u_1\right) - Z_T^{1/2}\left(v_2, w_2, u_2\right)\right|^2 \\ &\leq \frac{1}{4\sigma^2}\mathbf{E}_* \int_0^T \Bigg[\frac{\left(v_1 - v_2\right) \mathbb{1}_{\{X_t < \vartheta_3\}}}{\sqrt{T}} + \frac{\left(w_1 - w_2\right) \mathbb{1}_{\{X_t \geq \vartheta_3\}}}{\sqrt{T}} \\ &\quad + \left(\vartheta_1 - \vartheta_2 + \frac{v_2 - v_1 - w_2 + w_1}{\sqrt{T}}\right) \mathbb{1}_{\left\{\vartheta_3 + \frac{u_1}{T} < X_t < \vartheta_3 + \frac{u_2}{T}\right\}} \Bigg]^2 X_t^2 \mathrm{d}t \\ &\leq C_1\left(v_1 - v_2\right)^2 + C_2\left(w_1 - w_2\right)^2 + C_3\left|u_2 - u_1\right|. \end{aligned}$$

In the case of MLE this estimate is not sufficient because the condition $d > 3$ is not fulfilled. We slightly modify the proof of (28). Let us denote $\boldsymbol{u} = \left(v, w, u\right)$ and put

$$V_T = \left(\frac{Z_T\,(v_2, w_2, u_2)}{Z_T\,(v_1, w_1, u_1)} \right)^{\frac{1}{32}} .$$

Then

$$\begin{aligned}\mathbf{E}_\vartheta \left| Z_T^{\frac{1}{32}}\,(v_1, w_1, u_1) - Z_T^{\frac{1}{32}}\,(v_2, w_2, u_2) \right|^8 &= \mathbf{E}_\vartheta\, Z_T^{\frac{1}{4}}\,(v_2, w_2, u_2)\,|1 - V_T|^8 \\ \leq \left(\mathbf{E}_\vartheta\, Z_T^{\frac{1}{2}}\,(v_2, w_2, u_2) \right)^{\frac{1}{2}} \left(\mathbf{E}_\vartheta\, |1 - V_T|^{16} \right)^{\frac{1}{2}} &\leq \left(\mathbf{E}_\vartheta\, |1 - V_T|^{16} \right)^{\frac{1}{2}} .\end{aligned}$$

The process $V_t, 0 \leq t \leq T$ by Itô formula admits the representation

$$V_T = 1 - a \int_0^T V_t\,(\Delta S\,(X_t))^2\,\mathrm{d}t + b \int_0^T V_t\,(\Delta S\,(X_t))\,\mathrm{d}W_t$$

with corresponding constants $a > 0$ and $b > 0$ and $\Delta S\,(X_t) \equiv \Delta S_t$ is the difference of two trend coefficients. Hence

$$\begin{aligned}\mathbf{E}_\vartheta\, |1 - V_T|^{16} &\leq A\mathbf{E}_\vartheta \left(\int_0^T V_t\,(\Delta S_t)^2\,\mathrm{d}t \right)^{16} + B\mathbf{E}_\vartheta \left(\int_0^T V_t\,(\Delta S_t)\,\mathrm{d}t \right)^{16} \\ &\leq A\mathbf{E}_\vartheta \left(\int_0^T V_t\,(\Delta S_t)^2\,\mathrm{d}t \right)^{16} + C\mathbf{E}_\vartheta \left(\int_0^T V_t^2\,(\Delta S_t)^2\,\mathrm{d}t \right)^{8} .\end{aligned}$$

Further

$$\begin{aligned}\mathbf{E}_\vartheta \left(\int_0^T V_t^2\,(\Delta S_t)^2\,\mathrm{d}t \right)^8 &\leq \mathbf{E}_\vartheta \sup_{0\leq t\leq T} V_t^{16} \left(\int_0^T (\Delta S_t)^2\,\mathrm{d}t \right)^8 \\ &\leq \left(\mathbf{E}_\vartheta \sup_{0\leq t\leq T} V_t^{24} \right)^{\frac{2}{3}} \left(\mathbf{E}_\vartheta \left(\int_0^T (\Delta S_t)^2\,\mathrm{d}t \right)^{24} \right)^{\frac{1}{3}} \\ &\leq \left(\mathbf{E}_\vartheta \left(\int_0^T (\Delta S_t)^2\,\mathrm{d}t \right)^{24} \right)^{\frac{1}{3}}\end{aligned}$$

because $\mathbf{E}_\vartheta \sup_{0\leq t\leq T} V_t^{24} \leq 1$. Now with the help of (29) we can write

$$\begin{aligned}\mathbf{E}_\vartheta \left(\int_0^T (\Delta S\,(X_t))^2\,\mathrm{d}t \right)^{24} &= \mathbf{E}_\vartheta \left(T \int_{-\infty}^{\infty} (\Delta S\,(x))^2\, f_T^\circ\,(x)\,\mathrm{d}x \right)^{24} \\ &\leq C_1\,(v_2 - v_1)^{48} + C_2\,(w_2 - w_1)^{48} + C_3\,(u_2 - u_1)^{24} .\end{aligned}$$

After substitution of these estimates we obtain

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\vartheta}}\left|Z_T^{\frac{1}{32}}\left(v_1,w_1,u_1\right)-Z_T^{\frac{1}{32}}\left(v_2,w_2,u_2\right)\right|^8\\ \leq A\left|v_2-v_1\right|^8+B\left|w_2-w_1\right|^8+C\left|u_2-u_1\right|^4.\end{aligned}$$

Therefore for the values $|v_i|+|w_i|+|u_i|\leq R$ we have

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\vartheta}}&\left|Z_T^{\frac{1}{32}}\left(v_1,w_1,u_1\right)-Z_T^{\frac{1}{32}}\left(v_2,w_2,u_2\right)\right|^8\\ &\leq C\left(1+R^4\right)\left(\left|v_2-v_1\right|^4+\left|w_2-w_1\right|^4+\left|u_2-u_1\right|^4\right).\end{aligned}\tag{35}$$

Hence the condition **B** is fulfilled with $m=4$ and $d=4>3$ for the random field $Y_T(v,w,u)=Z_T^{\frac{1}{4}}(v,w,u)$.

To verify the condition **C** we follow the proof of Lemma 2.11 in Kutoyants (2004). We write $(u>0)$

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\vartheta}}J_T=&\frac{v^2}{\sigma^2T}\int_0^T\mathbf{E}_{\boldsymbol{\vartheta}}X_t^2\mathbb{1}_{\{X_t<\vartheta_3\}}\mathrm{d}t+\frac{w^2}{\sigma^2T}\int_0^T\mathbf{E}_{\boldsymbol{\vartheta}}X_t^2\mathbb{1}_{\{X_t\geq\vartheta_3\}}\mathrm{d}t\\ &+\left(\frac{\vartheta_2-\vartheta_1}{\sigma}+\frac{v-w}{\sigma\sqrt{T}}\right)^2\int_0^T\mathbf{E}_{\boldsymbol{\vartheta}}X_t^2\mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u}{T}\}}\mathrm{d}t\\ &+2\frac{w}{\sqrt{T}}\left(\frac{\vartheta_2-\vartheta_1}{\sigma}+\frac{v-w}{\sigma\sqrt{T}}\right)\int_0^T\mathbf{E}_{\boldsymbol{\vartheta}}X_t^2\mathbb{1}_{\{\vartheta_3<X_t\leq\vartheta_3+\frac{u}{T}\}}\mathrm{d}t.\end{aligned}$$

Note that

$$0<\kappa\equiv\frac{\alpha_2-\beta_1}{\sigma}<\frac{1}{\sigma}\left|\vartheta_2+\frac{w}{\sqrt{T}}-\left(\vartheta_1+\frac{v}{\sqrt{T}}\right)\right|<\frac{\beta_2-\alpha_1}{\sigma}\equiv K.$$

Hence

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\vartheta}}J_T\geq&\frac{v^2}{\sigma^2}\mathbf{E}_{\boldsymbol{\vartheta}}\xi^2\mathbb{1}_{\{\xi<\vartheta_3\}}+\frac{w^2}{\sigma^2}\mathbf{E}_{\boldsymbol{\vartheta}}\xi^2\mathbb{1}_{\{\xi\geq\vartheta_3\}}+\kappa^2T\int_{\vartheta_3}^{\vartheta_3+\frac{u}{T}}x^2f(\boldsymbol{\vartheta},x)\,\mathrm{d}x\\ &-2\frac{|w|}{\sqrt{T}}KT\int_{\vartheta_3}^{\vartheta_3+\frac{u}{T}}x^2f(\boldsymbol{\vartheta},x)\,\mathrm{d}x.\end{aligned}$$

Let us put $\delta=\kappa^2/4K$, then for $\frac{|v|}{\sqrt{T}}+\frac{|w|}{\sqrt{T}}+\frac{|u|}{T}\leq\delta$ we have

$$\mathbf{E}_{\boldsymbol{\vartheta}}J_T\geq v^2\mathrm{I}_1+w^2\mathrm{I}_2+|u|\,\tfrac{\kappa^2}{2}\alpha_3^2\inf_{\alpha_3<x\leq\beta_3}f(\boldsymbol{\vartheta},x),\tag{36}$$

and for the vector $\boldsymbol{h} = (h_1, h_2, h_3)$ with $h_1 = \frac{v}{\sqrt{T}}, h_2 = \frac{w}{\sqrt{T}}, h_3 = \frac{u}{T}$, and $\|\boldsymbol{h}\| \geq \delta$ we can write

$$\begin{aligned}\frac{\mathbf{E}_{\boldsymbol{\vartheta}} J_T}{T} &= h_1^2 \int_{-\infty}^{\vartheta_3} x^2 f(\boldsymbol{\vartheta}, x)\, \mathrm{d}x + h_2^2 \int_{\vartheta_3+h_3}^{\infty} x^2 f(\boldsymbol{\vartheta}, x)\, \mathrm{d}x \\ &\quad + (\vartheta_1 - \vartheta_2 + h_1)^2 \int_{\vartheta_3}^{\vartheta_3+h_3} x^2 f(\boldsymbol{\vartheta}, x)\, \mathrm{d}x \\ &\geq h_1^2 \int_{-\infty}^{\vartheta_3} x^2 f(\boldsymbol{\vartheta}, x)\, \mathrm{d}x + h_2^2 \int_{\beta_3}^{\infty} x^2 f(\boldsymbol{\vartheta}, x)\, \mathrm{d}x \\ &\quad + (\alpha_2 - \beta_1)^2 \int_{\vartheta_3}^{\vartheta_3+h_3} x^2 f(\boldsymbol{\vartheta}, x)\, \mathrm{d}x > \kappa_1 > 0. \end{aligned} \tag{37}$$

Here we used the representation

$$\begin{aligned}&-\vartheta_1 \mathbb{1}_{\{x<\vartheta_3\}} - \vartheta_2 \mathbb{1}_{\{x\geq\vartheta_3\}} + (\vartheta_1 + h_1)\, \mathbb{1}_{\{x<\vartheta_3+h_3\}} - (\vartheta_2 + h_2)\, \mathbb{1}_{\{x\geq\vartheta_3+h_3\}} \\ &= h_1 \mathbb{1}_{\{x<\vartheta_3\}} + h_2 \mathbb{1}_{\{x\geq\vartheta_3+h_3\}} + (\vartheta_1 - \vartheta_2 + h_1)\, \mathbb{1}_{\{\vartheta_3<x\leq\vartheta_3+h_3\}}.\end{aligned}$$

Now having (36) and (37), we can follow the proof of Lemma 2.11 in Kutoyants (2004) and obtain the estimate (24). □

5.3 Proof of (9)

The properties of modified (simplified) estimators defined by the equalities (8) will be proved if we verify the law of large numbers (9). For any $\varepsilon > 0$ using the consistency (7), we can write

$$\begin{aligned}&\mathbf{P}_{\boldsymbol{\vartheta}} \left\{ \left| \frac{1}{\sigma^2 T} \int_{\sqrt{T}}^{T} X_t^2 \mathbb{1}_{\left\{X_t<\hat{\vartheta}^*_{3,\sqrt{T}}\right\}} \mathrm{d}t - \mathrm{I}_1 \right| > \varepsilon \right\} \leq \mathbf{P}_{\boldsymbol{\vartheta}} \left\{ \left| \hat{\vartheta}^*_{3,\sqrt{T}} - \vartheta_3 \right| \geq T^{-b} \right\} \\ &\quad + \mathbf{P}_{\boldsymbol{\vartheta}} \left\{ \left| \frac{1}{\sigma^2 T} \int_{\sqrt{T}}^{T} X_t^2 \mathbb{1}_{\left\{X_t<\hat{\vartheta}^*_{3,\sqrt{T}}\right\}} \mathrm{d}t - \mathrm{I}_1 \right| > \varepsilon, \left| \hat{\vartheta}^*_{3,T} - \vartheta_3 \right| < T^{-b} \right\} \\ &\leq \mathbf{P}_{\boldsymbol{\vartheta}} \left\{ \sup_{|\theta-\vartheta_3|<T^{-b}} \left| \frac{1}{\sigma^2 T} \int_{\sqrt{T}}^{T} X_t^2 \mathbb{1}_{\{X_t<\theta\}} \mathrm{d}t - \mathrm{I}_1 \right| > \varepsilon \right\} + o(1).\end{aligned}$$

Further

$$\begin{aligned}&\sup_{|\theta-\vartheta_3|<T^{-b}} \left| \frac{1}{\sigma^2 T} \int_{\sqrt{T}}^{T} X_t^2 \mathbb{1}_{\{X_t<\theta\}} \mathrm{d}t - \mathrm{I}_1 \right| \\ &\quad = \sup_{|\theta-\vartheta_3|<T^{-b}} \left| \frac{1}{\sigma^2 T} \int_{0}^{T} X_t^2 \mathbb{1}_{\{X_t<\theta\}} \mathrm{d}t - \mathrm{I}_1 \right| + o(1)\end{aligned}$$

$$= \sup_{|\theta-\vartheta_3|<T^{-b}} \left| \int_{-\infty}^{\theta} \frac{x^2}{\sigma^2} f_T^\circ(x)\,\mathrm{d}x - \int_{-\infty}^{\vartheta_3} \frac{x^2}{\sigma^2} f(\boldsymbol{\vartheta},x)\,\mathrm{d}x \right| + o(1)$$

$$\leq \left| \int_{-\infty}^{\vartheta_3} \frac{x^2}{\sigma^2} \left[f_T^\circ(x) - f(\boldsymbol{\vartheta},x) \right] \mathrm{d}x \right| + \int_{\vartheta_3}^{\vartheta_3+T^{-b}} \frac{x^2}{\sigma^2} f_T^\circ(x)\,\mathrm{d}x + o(1).$$

Here $f_T^\circ(x) = \sigma^{-2} T^{-1} \Lambda_T(x)$ is the local time estimator of the invariant density. To finish the proof we just mention, that

$$\int_{-\infty}^{\vartheta_3} \frac{x^2}{\sigma^2} \left[f_T^\circ(x) - f(\boldsymbol{\vartheta},x) \right] \mathrm{d}x \longrightarrow 0$$

by the law of large numbers. □

6 Discussion

6.1 Goodness-of-fit testing

Suppose that the basic hypothesis ($\mathscr{H}_0$) is simple: the observations $X^T = (X_t, 0 \leq t \leq T)$ come from the Eq. (12) with known $\boldsymbol{\vartheta}_0$. There are several ways to construct the goodness-of-fit tests. It is possible, for example, to use the Cramér-von Mises and Kolmogorov–Smirnov type statistics proposed in Dachian and Kutoyants (2007) or in Negri and Nishiyama (2009). Let us discuss another approach developed in Kutoyants (2010) which is more close to classical statement. Introduce the empirical distribution function $\hat{F}_T(\cdot) = T^{-1} \int_0^T \mathbb{1}_{\{X_t<x\}}\,\mathrm{d}t$ and the corresponding C-vM statistics

$$\mathbb{V}_T^2 = T \int_{-\infty}^{\infty} H(\boldsymbol{\vartheta}_0,x) \left[\hat{F}_T(x) - F(\boldsymbol{\vartheta}_0,x) \right]^2 \mathrm{d}F(\boldsymbol{\vartheta}_0,x),$$

with weight function

$$H(\boldsymbol{\vartheta}_0,x) = \frac{\Psi'(\boldsymbol{\vartheta}_0,x)}{f(\boldsymbol{\vartheta}_0,x)\left[F(\boldsymbol{\vartheta}_0,x)-1\right]^2} M(\Psi(\boldsymbol{\vartheta}_0,x)).$$

where $M(\cdot)$ is some function providing the finitness of this integral and

$$\Psi(\boldsymbol{\vartheta}_0,x) = \int_{-\infty}^{x} \frac{F(\boldsymbol{\vartheta}_0,y)^2}{\sigma(y)^2 f_0(\boldsymbol{\vartheta}_0,y)} \mathrm{d}y + F(\boldsymbol{\vartheta}_0,x)^2 \int_x^\infty \left(\frac{F(\boldsymbol{\vartheta}_0,y)-1}{F(\boldsymbol{\vartheta}_0,x)-1} \right)^2 \frac{\mathrm{d}y}{\sigma(y)^2 f(\boldsymbol{\vartheta}_0,y)}.$$

It is shown that if $M(s) = e^{-s}$ then

$$\mathbb{V}_T^2(\boldsymbol{\vartheta}_0) \Longrightarrow \int_0^\infty W(s)^2\, e^{-s}\,\mathrm{d}s,$$

where $W(\cdot)$ is a Wiener process, i.e.; we have asymptotically distribution free test $\hat\psi_T = \mathbb{1}_{\{\mathbb{V}_T^2(\boldsymbol{\vartheta}_0)>r_\alpha\}}$ (see Kutoyants (2010)). The threshold r_α, of course, is solution of the following equation $\mathbf{P}\left\{\int_0^\infty W(s)^2 e^{-s}\,\mathrm{d}s > r_\alpha\right\} = \alpha$. The similar result can be proved for the large class of functions $M(\cdot)$ satisfying the obvious conditions.

In the case of composite basic hypothesis it is supposed that the observed process satisfies the equation (12), but the value of $\boldsymbol{\vartheta}$ is unknown. Then we can use the same statistic with parameter replaced by some estimator(MLE or BE) and show that the test $\hat\psi_T = \mathbb{1}_{\left\{\mathbb{V}_T^2\left(\hat{\boldsymbol{\vartheta}}_T\right)>r_\alpha\right\}}$ with the same r_α is asymptotically distribution-free (see Kutoyants 2010 for details).

6.2 Conclusion

The studied estimators have some particularities which are typical for singular estimation problems. Let us mention some of them.

1. The rate of convergence is T and not $\sqrt{T}$ as in regular models.
2. The asymptotically efficient estimators are BE and not MLE.
3. The likelihood ratios (LR) converge to the exponential functional of Wiener process (not LAN).
4. The estimation problem is more robust than in regular case and admits the consistent estimation even for wrong models.
5. The one-step MLE does not exist but it is possible to “localize” the problem by the first $\sqrt{T}$ observations and to use the narrow windows for observations.
6. In multidimensional case the limit LR is a product of one-dimensional independent LRs.
7. The goodness-of-fit test even with parametric basic hypothesis is asymptotically distribution-free.

Of course, there are a lot of open problems. For example, it can be interesting to study the properties of estimators for multidimensional threshold diffusion processes, partially observed linear switching systems or to use the multithreshold model for approximation of some trend coefficients.

Acknowledgments I would like to thank the associate editor and two referees for the useful comments.

References

Chan, K. S. (1993). Consistency and limiting distribution of the LSE of a TAR. *The Annals of Statistics*, *21*, 520–533.

Chan, N. H., Kutoyants, Yu. A. (2008). On parameter estimation of threshold autoregressive models (arXiv:1003.3800) (submitted).

Dachian, S., Kutoyants, Yu. A. (2007). On the goodness-of-fit tests for some continuous time processes. In F. Vonta et al. (Eds.), *Statistical models and methods for biomedical and technical systems* (pp. 395–413). Boston: Birkhäuser.

Decamps, M., Goovaerts, M., Schoutens, W. (2006). Self exciting threshold interest rates models. *International Journal of Theoretical and Applied Finance*, *9*, 1093–1122.

Durret, R. (1996). *Stochastic calculus: A practical introduction*. Boca Raton: CRC Press.

Fan, J., Yao, Q. (2003). *Nonlinear time series: Nonparametric and parametric methods*. New York: Springer.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, *68*, 575–603.
Ibragimov, I. A., Khasminskii, R. Z. (1981). *Statistical estimation*. New York: Springer.
Koul, H. L., Qianb, L., Surgailis, D. (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Processes and Applications*, *103*, 123–154.
Küchler, U., Kutoyants, Yu. A. (2000). Delay estimation for some stationary diffusion-type processes. *Scandinavian Journal of Statistics*, *27*(3), 405–414.
Kutoyants, Yu. A. (2004). *Statistical inference for ergodic diffusion processes*. London: Springer.
Kutoyants, Yu. A. (2010). On the goodness-of-fit testing for ergodic diffusion processes. *Journal of Nonparametric Statistics*, *22*(4), 529–543.
Lehmann, E. L., Romano, J. P. (2005). *Testing statistical hypotheses* (3rd ed.). New York: Springer.
Liptser, R. S., Shiryayev, A. N. (2001). *Statistics of random processes. I* (2nd ed.). New York: Springer.
Negri, I., Nishiyama, Y. (2009). Goodness of fit test for ergodic diffusion processes. *Annals of the Institute of Statistical Mathematics*, *61*(4), 919–928.
Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of American Statistical Association*, *53*, 873–880.
Revuz, D., Yor, M. (1991). *Continuous martingales and brownian motion*. New York: Springer.
Rubin, H., Song, K.-S. (1995). Exact computation of the asymptotic efficiency of maximum likelihood estimators of a discontinuous signal in a Gaussian white noise. *The Annals of Statistics*, *23*, 732–739.
Shreve, S. E. (2004). *Stochastic calculus for finance II: Continuous-time models*. New York: Springer.
Terent'yev, A. S. (1968). Probability distribution of a time location of an absolute maximum at the output of a synchronized filter. *Radioengineering and Electronics*, *13*(4), 652–657.
Tong, H. (1990). *Non-linear time series: A dynamical systems approach*. Oxford: Oxford University Press.
Yoshida, N. (2009). Polynomial type large deviation inequality and its applications. *Annals of the Institute of Statistical Mathematics* (to appear).