Stat Inference Stoch Process (2019) 22:233–261
https://doi.org/10.1007/s11203-018-9176-x

# Empirical $L^2$-distance test statistics for ergodic diffusions

**A. De Gregorio[1] · S. M. Iacus[2]**

Received: 11 May 2017 / Accepted: 10 February 2018 / Published online: 21 February 2018


**Abstract** The aim of this paper is to introduce a new type of test statistic for simple null hypothesis on one-dimensional ergodic diffusion processes sampled at discrete times. We deal with a quasi-likelihood approach for stochastic differential equations (i.e. local gaussian approximation of the transition functions) and define a test statistic by means of the empirical $L^2$-distance between quasi-likelihoods. We prove that the introduced test statistic is asymptotically distribution free; namely it weakly converges to a $\chi^2$ random variable. Furthermore, we study the power under local alternatives of the parametric test. We show by the Monte Carlo analysis that, in the small sample case, the introduced test seems to perform better than other tests proposed in literature.

**Keywords** Asymptotic distribution free test · Local alternatives · Maximum-likelihood type estimator · Discrete observations · Quasi-likelihood function · Stochastic differential equation

## 1 Introduction

Let $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered complete probability space. Let us consider a 1-dimensional processes $X := (X_t)_{t\geq 0}$ solution to the following stochastic differential equation

$$\mathrm{d}X_t = b(\alpha, X_t)\mathrm{d}t + \sigma(\beta, X_t)\mathrm{d}W_t, \quad X_0 = x_0, \tag{1.1}$$

✉ A. De Gregorio
alessandro.degregorio@uniroma1.it

S. M. Iacus
stefano.iacus@unimi.it

1 Department of Statistical Sciences, "Sapienza" University of Rome, P.le Aldo Moro, 5, 00185 Rome, Italy

2 Department of Economics, Management and Quantitative Methods, University of Milan, Via Conservatorio 7, 20122 Milan, Italy

where $x_0$ is a deterministic initial value. We assume that $b : \Theta_\alpha \times \mathbb{R} \to \mathbb{R}, \sigma : \Theta_\beta \times \mathbb{R} \to \mathbb{R}$ are Borel known functions (up to $\alpha$ and $\beta$) and $(W_t)_{t\geq 0}$ is a one-dimensional standard $\mathcal{F}_t$-Brownian motion. Furthermore, $\alpha \in \Theta_\alpha \subset \mathbb{R}^{m_1}$, $\beta \in \Theta_\beta \subset \mathbb{R}^{m_2}$, $m_1, m_2 \in \mathbb{N}$, are unknown parameters and $\theta = (\alpha, \beta) \in \Theta := \Theta_\alpha \times \Theta_\beta$, where $\Theta_\alpha$ and $\Theta_\beta$ are compact convex sets. We denote by $\theta_0 := (\alpha_0, \beta_0)$ the true value of $\theta$ and assume that $\theta_0 \in \mathrm{Int}(\Theta)$.

The sample path of $X$ is observed only at $n+1$ equidistant discrete times $t_i^n$, such that $t_i^n - t_{i-1}^n = \Delta_n < \infty$ for $i = 1, \dots, n$, (with $t_0^n = 0$). Therefore the data, denoted by $(X_{t_i^n})_{0\leq i\leq n}$, are the discrete observations of the sample path of $X$. Let $p$ be an integer with $p \geq 2$. The asymptotic scheme adopted in this paper is the following: $n\Delta_n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n^p \to 0$ as $n \to \infty$. This scheme is called rapidly increasing design, i.e. the number of observations grows over time but no so fast.

This setting is useful, for instance, in the analysis of financial time series. In mathematical finance and econometric theory, diffusion processes described by the stochastic differential equations (1.1) play a central role. Indeed, they have been used to model the behavior of stock prices, exchange rates and interest rates. The underlying stochastic evolution of the financial assets can be thought continuous in time, although the data are always recorded at discrete instants (e.g. weekly, daily or each minute). For these reasons, the estimation problems for discretely observed stochastic differential equations have been tackled by many authors with different approaches (see, for instance, Florens-Zmirou 1989; Yoshida 1992; Genon-Catalot and Jacod 1993; Bibby and Sørensen 1995; Kessler 1997; Kessler and Sørensen 1999; Aït-Sahalia 2002; Gobet 2002; Jacod 2006; Aït-Sahalia 2008; De Gregorio and Iacus 2008; Phillips and Yu 2009; Yoshida 2011; Uchida and Yoshida 2012; Li 2013; Uchida and Yoshida 2014; Kamatani and Uchida 2015). For clustering time series arising from discrete observations of diffusion processes De Gregorio and Iacus (2010) proposed a new dissimilarity measure based on the $L^1$ distance between the Markov operators. The change-point problem in the diffusion term of a stochastic differential equation has been considered in De Gregorio and Iacus (2008) and Iacus and Yoshida (2012). In Iacus et al. (2009), the authors faced the estimation problem for hidden diffusion processes observed at discrete times. An adaptive Lasso-type estimator is proposed in De Gregorio and Iacus (2012). For the simulation and the practical implementation of the statistical inference for stochastic differential equations see Iacus (2008, 2011) and Iacus and Yoshida (2017).

We also recall that the statistical inference for continuously observed ergodic diffusions is a well-developed research topic; on this point the reader can consult Kutoyants (2004).

The main object of interest of the present paper is the problem of testing parametric hypotheses for diffusion processes from discrete observations. This research topic is less developed in literature. It is well-known that for testing two simple alternative hypotheses, the Neyman-Pearson lemma provides a procedure based on the likelihood ratio which leads to the uniformly most powerful test. In the other cases uniformly most powerful tests do not exist and for this reason the research of new criteria is justified.

For discretely observed stochastic differential equations, Kitagawa and Uchida (2014) introduced and studied the asymptotic behavior of three kinds of test statistics: likelihood ratio type test statistic, Wald type test statistic and Rao's score type test statistic.

Another possible approach is based on the divergences. Indeed, several statistical divergence measures (which are not necessarily a metric) and distances have been introduced to decide if two probability distributions are close or far. The main goal of this metric is to make "easy to distinguish" between a pair of distributions which are far from each other than between those which are closer. These tools have been used for testing hypotheses in parametric models. The reader can consult on this point, for example, Morales et al. (1997) and Pardo (2006). For stochastic differential equations sampled at discrete times, De Gregorio and Iacus (2013) introduced a family of test statistics (for $p = 2$ and $n\Delta_n^2 \to 0$) based on empirical $\phi$-divergences.

We consider the following hypotheses testing problem concerning the vector parameter $\theta$

$$H_0 : \theta = \theta_0, \quad \text{vs} \quad H_1 : \theta \neq \theta_0,$$

and assume that $X$ is observed at discrete times; that is the data $(X_{t_i^n})_{0\le i\le n}$ are available. In this work we study different test statistics with respect to those used in De Gregorio and Iacus (2013) and Kitagawa and Uchida (2014). Indeed, the purpose of this paper is to propose a methodology based on a suitable "distance" between the approximated transition functions. This idea follows from the observation that for continuously observed sample paths of (1.1), we could define the $L^2$-distance between the continuous log-likelihood. Clearly this approach is not useful in our framework and then, similarly to the aforementioned papers, we consider the local gaussian approximation of the transition density of the process $X$ from $X_{t_{i-1}}$ to $X_{t_i}$. In other words, we resort the quasi-likelihood function introduced in Kessler (1997), which is defined by means of an approximation with higher order correction terms to relax the condition of convergence of $\Delta_n$ to zero. Therefore, let $\mathtt{l}_{p,i}(\theta), \theta \in \Theta$, be the approximated log-transition function from $X_{t_{i-1}}$ to $X_{t_i}$ representing the parametric model (1.1). We deal with

$$\mathbb{D}_{p,n}(\theta_1, \theta_2) := \frac{1}{n}\sum_{i=1}^{n}[\mathtt{l}_{p,i}(\theta_1) - \mathtt{l}_{p,i}(\theta_2)]^2, \quad \theta_1, \theta_2 \in \Theta,$$

which can be interpreted as the empirical $L^2$-distance between two loglikelihoods. If $\hat{\theta}_{p,n}$ is the maximum quasi-likelihood estimator introduced in Kessler (1997), we are able to prove that, under $H_0$, the test statistic

$$T_{p,n}(\hat{\theta}_{p,n}, \theta_0) := n\mathbb{D}_{p,n}(\hat{\theta}_{p,n}, \theta_0)$$

is asymptotically distribution free; i.e. it converges in distribution to a chi squared random variable. Furthermore, we study the power function of the test under local alternatives.

The paper is organized as follows. Section 2 contains the notations and the assumptions of the paper. The contrast function arising from the quasi-likelihood approach is briefly discussed in Sect. 3. In the same section we define the maximum quasi-likelihood estimator and recall its main asymptotic properties. In Sect. 4 we introduce and study a test statistic for the hypotheses problem $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. The proposed new test statistic shares the same asymptotic properties of the other test statistics presented in the literature. Therefore, to justify its use in practice among its competitors, a numerical study is included in Sect. 5 which contains a comparison of several test statistics in the "small sample" case, i.e., when the asymptotic conditions are not met. Our numerical analysis shows that, at least for $p = 2$, the performance of $T_{2,n}$ is very good. The proofs are collected in Sect. 6.

It is worth to point out that for the sake of simplicity in this paper a 1-dimensional diffusion is treated. Nevertheless, it is possible to extend our methodology to the multidimensional stochastic differential equations setting.

## 2 Notations and assumptions

Throughout this paper, we will use the following notation.

- $\theta := (\alpha, \beta)$ and $\alpha_0$, $\beta_0$ and $\theta_0$ denote the true values of $\alpha$, $\beta$ and $\theta$ respectively.

- $c(\beta, x) = \sigma^2(\beta, x)$.
- $C$ is a positive constant. If $C$ depends on a fixed quantity, for instance an integer $k$, we may write $C_k$.
- $\partial_{\alpha_h} := \frac{\partial}{\partial \alpha_h}, \partial_{\beta_k} := \frac{\partial}{\partial \beta_k}, \partial^2_{\alpha_h \alpha_k} := \frac{\partial^2}{\partial \alpha_h \partial \alpha_k}, h, k = 1, \ldots, m_1, \partial^2_{\beta_h \beta_k} := \frac{\partial^2}{\partial \beta_h \partial \beta_k}, h, k = 1, \ldots, m_2, \partial^2_{\alpha_h \beta_k} := \frac{\partial^2}{\partial \alpha_h \partial \beta_k}, h = 1, \ldots, m_1, k = 1, \ldots, m_2, \partial_\theta := (\partial_\alpha, \partial_\beta)'$, where $\partial_\alpha := (\partial_{\alpha_1}, \ldots, \partial_{\alpha_{m_1}})'$ and $\partial_\beta := (\partial_{\beta_1}, \ldots, \partial_{\beta_{m_2}})'$, $\partial^2_\theta := [\partial^2_{\alpha_j \beta_k}]_{h=1,\ldots,m_1,k=1,\ldots,m_2}$.
- If $f : \Theta \times \mathbb{R} \to \mathbb{R}$, we denote by $f_{i-1}(\theta)$ the value $f(\theta, X_{t^n_{i-1}})$; for instance $c(\beta, X_{t^n_{i-1}}) = c_{i-1}(\beta)$.
- For $0 \leq i \leq n$, $t^n_i := i\Delta_n$ and $\mathcal{G}^n_i := \sigma(W_s, s \leq t^n_i)$.
- The random sample is given by $\mathbf{X}_n := (X_{t^n_i})_{0 \leq i \leq n}$ and $X_i := X_{t^n_i}$.
- The probability law of (1.1) is denoted by $P_\theta$ and $E^{i-1}_\theta[\cdot] := E_\theta[\cdot|\mathcal{G}^n_{i-1}]$. We set $P_0 := P_{\theta_0}$ and $E^{i-1}_0[\cdot] := E^{i-1}_{\theta_0}[\cdot]$.
- $\xrightarrow[n\to\infty]{P_\theta}$ and $\xrightarrow[n\to\infty]{d}$ stand for the convergence in probability and in distribution, respectively.
- Let $F_n : \Theta \times \mathbb{R}^n \to \mathbb{R}$ and $F : \Theta \to \mathbb{R}$; "$F_n(\theta, \mathbf{X}_n) \xrightarrow[n\to\infty]{P_\theta} F(\theta)$ uniformly in $\theta$" stands for

$$\sup_{\theta \in \Theta} |F_n(\theta, \mathbf{X}_n) - F(\theta)| \xrightarrow[n\to\infty]{P_\theta} 0.$$

  Furthermore, if $F_n(\theta, \mathbf{X}_n) \xrightarrow[n\to\infty]{P_\theta} 0$ uniformly in $\theta$ we set

$$F_n(\theta, \mathbf{X}_n) = \mathbf{o}_{P_\theta}(1).$$

- Let $u_n$ be a $\mathbb{R}$-valued sequence. We indicate by $R$ a function $\Theta \times \mathbb{R}^2 \to \mathbb{R}$ for which there exists a constant $C$ such that

$$R(\theta, u_n, x) \leq u_n C(1 + |x|)^C, \quad \text{for all}\, \theta \in \Theta, x \in \mathbb{R}^2, n \in \mathbb{N}.$$

- For a $m \times n$ matrix $A$, $||A||^2 = \mathrm{tr}(AA') = \sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2$ and $I_m$ stands for the identity matrix of size $m$.

Let $C^{k,h}_\uparrow(\mathbb{R} \times \Theta; \mathbb{R})$ be the space of all functions $f$ such that:

(i) $f(\theta, x)$ is a $\mathbb{R}$-valued function on $\Theta \times \mathbb{R}$;
(ii) $f(\theta, x)$ is continuously differentiable with respect to $x$ up to order $k \geq 1$ for all $\theta$; these $x$-derivatives up to order $k$ are of polynomial growth in $x$, uniformly in $\theta$;
(iii) $f(\theta, x)$ and all $x$-derivatives up to order $k \geq 1$, are $h \geq 1$ times continuously differentiable with respect to $\theta$ for all $x \in \mathbb{R}$. Moreover, these derivatives up to the $h$-th order with respect to $\theta$ are of polynomial growth in $x$, uniformly in $\theta$.

We need some standard assumptions on the regularity of the process $X$.

$A_1$ (Existence and uniqueness) There exists a constant $C$ such that

$$\sup_{\alpha \in \Theta_\alpha} |b(\alpha, x) - b(\alpha, y)| + \sup_{\beta \in \Theta_\beta} |\sigma(\beta, x) - \sigma(\beta, y)| \leq C|x - y|.$$

$A_2$ (Ergodicity) The process $X$ is ergodic for $\theta = \theta_0$ with invariant probability measure $\pi_0(\mathrm{d}x)$. Thus

$$\frac{1}{T}\int_0^T f(X_t)\mathrm{d}t \xrightarrow[T\to\infty]{P_\theta} \int f(x)\pi_0(\mathrm{d}x),$$

where $f \in L^1(\pi_0)$. Furthermore, we assume that $\pi_0$ admits all moments finite.

$A_3$ $\inf_{x,\beta} \sigma(\beta, x) > 0$.
$A_4$ (Moments) For all $q \geq 0$, $\sup_t E|X_t|^q < \infty$.
$A_5$ $[k]$ (Smoothness) $b \in C_{\uparrow}^{k,3}(\Theta_\alpha \times \mathbb{R}, \mathbb{R})$ and $\sigma \in C_{\uparrow}^{k,3}(\Theta_\beta \times \mathbb{R}, \mathbb{R})$.
$A_6$ (Identifiability) If the coefficients $b(\alpha, x) = b(\alpha_0, x)$ and $\sigma(\beta, x) = \sigma(\beta_0, x)$ for all $x$ ($\pi_0$-almost surely), then $\alpha = \alpha_0$ and $\beta = \beta_0$.

Let $L_\theta$ the infinitesimal generator of $X$ with domain given by $C^2(\mathbb{R})$ (the space of the twice continuously differentiable function on $\mathbb{R}$); that is if $f \in C^2(\mathbb{R})$

$$L_\theta f(x) := b(\alpha, x)\frac{\partial f}{\partial x}(x) + \frac{c(\beta, x)}{2}\frac{\partial^2 f}{\partial x^2}(x), \quad L_0 := L_{\theta_0}.$$

Under the assumption $A_5[2(j-1)]$ we can define $L_\theta^j := L_\theta \circ L_\theta^{j-1}$ with domain $C^{2j}(\mathbb{R})$ and $L_\theta^0 =$Id.

We conclude this section with some well-known examples of ergodic diffusion processes belonging to the class (1.1):

- the Ornstein–Uhlenbeck or Vasicek model is the unique solution to

$$\mathrm{d}X_t = \alpha_1(\alpha_2 - X_t)\mathrm{d}t + \beta_1 \mathrm{d}W_t, \quad X_0 = x_0, \tag{2.1}$$

  where $b(\alpha_1, \alpha_2, x) = \alpha_1(\alpha_2 - x)$ and $\sigma(\beta_1, x) = \beta_1$ with $\alpha_1, \alpha_2 \in \mathbb{R}$ and $\beta_1 > 0$. This stochastic process is a Gaussian process and it is often used in finance where $\beta_1$ is the volatility, $\alpha_2$ is the long-run equilibrium of the model and $\alpha_1$ is the speed of mean reversion. For $\alpha_1 > 0$ the Vasicek process is ergodic with invariant law $\pi_0$ given by a Gaussian law with mean $\alpha_2$ and variance $\frac{\beta_1^2}{2\alpha_1}$. It is easy to check that all the conditions $A_1 - A_6$ fulfill;
- the Cox–Ingersoll–Ross (CIR) process is the solution to

$$\mathrm{d}X_t = \alpha_1(\alpha_2 - X_t)\mathrm{d}t + \beta_1\sqrt{X_t}\mathrm{d}W_t, \quad X_0 = x_0 > 0, \tag{2.2}$$

  where $b(\alpha_1, \alpha_2, x) = \alpha_1(\alpha_2 - x)$ and $\sigma(\beta_1, x) = \beta_1\sqrt{x}$ with $\alpha_1, \alpha_2, \beta_1 > 0$. If $2\alpha_1\alpha_2 > \beta_1^2$ the process is strictly positive, otherwise non negative. This model has a conditional density given by the non central $\chi^2$ distribution. The CIR process is useful in the description of short-term interest rates and admits invariant law $\pi_0$ given by a Gamma distribution with shape parameter $\frac{2\alpha_1\alpha_2}{\beta_1^2}$ and scale parameter $\frac{\beta_1^2}{2\alpha_1}$. If (2.2) is strictly positive, we can prove that the above assumptions hold true.

## 3 Preliminaries on the quasi-likelihood function

We briefly recall the quasi-likelihood function introduced by Kessler (1997) based on the Itô-Taylor expansion. The main problem in the statistical analysis of the diffusion process $X$ is that its transition density is in general unknown and then the likelihood function is unknown as well. To overcome this difficulty one can discretizes the sample path of $X$ by means of Euler-Maruyama's scheme; namely

$$X_i - X_{i-1} = \int_{t_{i-1}^n}^{t_i^n} b(\alpha, X_s)\mathrm{d}s + \int_{t_{i-1}^n}^{t_i^n} \sigma(\beta, X_s)\mathrm{d}W_s \simeq b_{i-1}(\alpha)\Delta_n + \sigma_{i-1}(\beta)(W_{t_i^n} - W_{t_{i-1}^n}). \tag{3.1}$$

Hence (3.1) leads to consider a local-Gaussian approximation to the transition density; that is

$$\mathcal{L}(X_i|X_{i-1}) \simeq N(X_{i-1} + b_{i-1}(\alpha)\Delta_n, c_{i-1}(\beta)\Delta_n)$$

and the approximated log-likelihood function of the random sample $\mathbf{X}_n$, called (negative) quasi-log-likelihood function, becomes

$$l_n(\theta) := \frac{1}{2}\sum_{i=1}^{n}\left\{\frac{(X_i - X_{i-1} - b_{i-1}(\alpha)\Delta_n)^2}{c_{i-1}(\beta)\Delta_n} + \log c_{i-1}(\beta)\right\}. \tag{3.2}$$

This approach suggests to consider the mean and the variance of the transition density of $X$; that is

$$\mathrm{m}(\theta, X_{i-1}) := E_\theta[X_i|X_{i-1}], \quad \mathrm{m}_2(\theta, X_{i-1}) := E_\theta[(X_i - \mathrm{m}(\theta, X_{i-1}))^2|X_{i-1}], \tag{3.3}$$

and assume

$$\mathcal{L}(X_i|X_{i-1}) \simeq N(\mathrm{m}(\theta, X_{i-1}), \mathrm{m}_2(\theta, X_{i-1})).$$

Thus we can consider as contrast function the following one

$$\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{(X_i - \mathrm{m}(\theta, X_{i-1}))^2}{\mathrm{m}_2(\theta, X_{i-1})} + \log \mathrm{m}_2(\theta, X_{i-1})\right\}. \tag{3.4}$$

Nevertheless, (3.4) does not have a closed form because $\mathrm{m}(\theta, X_{i-1})$ and $\mathrm{m}_2(\theta, X_{i-1})$ are unknown. Therefore we substitute in (3.4) closed approximations of m and $\mathrm{m}_2$ based on the Itô-Taylor expansion.

Let $f(y) := y$, for $l \geq 0$, under the assumption $A_5[2(l-1)]$, we have the following approximation (see Lemma 1, Kessler 1997)

$$\mathrm{m}(\theta, X_{i-1}) = r_l(\Delta_n, X_{i-1}, \theta) + R(\theta, \Delta_n^{l+1}, X_{i-1}) \tag{3.5}$$

where

$$r_l(\Delta_n, X_{i-1}, \theta) := \sum_{i=0}^{l} \frac{\Delta_n^i}{i!} L_\theta^i f(x).$$

Now let us consider the function $(y - r_l(\Delta_n, X_{i-1}, \theta))^2$, which is for fixed $x$, $y$ and $\theta$ a polynomial in $\Delta_n$ of degree $2l$. We indicate by $\overline{g}_{\Delta_n,x,\theta,l}(y)$ the sum of its first terms up to degree $l$; that is $\overline{g}_{\Delta_n,x,\theta,l}(y) = \sum_{j=0}^{l} \Delta_n^j \overline{g}_{x,\theta}^j(y)$ where

$$\overline{g}_{x,\theta}^0(y) = (y-x)^2 \tag{3.6}$$

$$\overline{g}_{x,\theta}^1(y) = -2(y-x)L_\theta f(x) \tag{3.7}$$

$$\overline{g}_{x,\theta}^j(y) = -2(y-x)\frac{L_\theta^j f(x)}{j!} + \sum_{r,s\geq 1, r+s=j} \frac{L_\theta^r f(x)}{r!}\frac{L_\theta^s f(x)}{s!}, \quad 2 \leq j \leq l. \tag{3.8}$$

Under the assumption $A_5[2(l-1)]$, we have that $L_\theta^r \overline{g}_{x,\theta}^j(y)$ is well-defined for $r + j = l$ and we set

$$\Gamma_l(\Delta_n, x, \theta) := \sum_{j=0}^{l} \Delta_n^j \sum_{r=0}^{l-j} \frac{\Delta_n^r}{r!} L_\theta^r \overline{g}_{x,\theta}^j(x) := \sum_{j=0}^{l} \Delta_n^j \gamma_j(\theta, x), \tag{3.9}$$

where $\gamma_j(\theta, x)$ are the coefficients of $\Delta_n^j$. Therefore by (3.6)–(3.9), we obtain, for instance,

$$\begin{aligned}
\gamma_0(\theta, x) &= L_\theta^0 \overline{g}_{x,\theta}^0(x) = 0 \\
\gamma_1(\theta, x) &= L_\theta \overline{g}_{x,\theta}^0(x) = c(\beta, x) \\
\gamma_2(\theta, x) &= \frac{L_\theta^2 \overline{g}_{x,\theta}^0}{2}(x) + L_\theta \overline{g}_{x,\theta}^1(x) + L_\theta^0 \overline{g}_{x,\theta}^2(x) \\
&= \frac{1}{2}\left[b(\alpha, x)\frac{\partial}{\partial y}c(\beta, x) + 2c(\beta, x)\frac{\partial}{\partial y}b(\alpha, x)\right] + \frac{c(\beta, x)}{4}\frac{\partial^2}{\partial y^2}c(\beta, x)
\end{aligned}$$

Let

$$\Gamma_l(\Delta_n, x, \theta) := \Delta_n c(\beta, x)[1 + \overline{\Gamma}_l(\Delta_n, x, \theta)]$$

where $\overline{\Gamma}_l(\Delta_n, x, \theta) := \frac{\sum_{j=2}^l \Delta_n^j \gamma_j(\theta, x)}{\Delta_n c(\beta, x)}$. For $l \geq 0$, under the assumption $A_5[2l]$(i), we have that (see Lemma 2, Kessler 1997)

$$\mathrm{m}_2(\theta, X_{i-1}) = \Delta_n c_{i-1}(\beta)[1 + \overline{\Gamma}_l(\Delta_n, X_{i-1}, \theta)] + R(\theta, \Delta_n^{l+1}, X_{i-1}). \tag{3.10}$$

It seems quite natural at this point to substitute (3.5) and (3.10) into the expression (3.4). Nevertheless, in order to avoid technical difficulties related to the control of denominator and logarithmic we consider a further expansion in $\Delta_n$ of $(1 + \overline{\Gamma}_l)^{-1}$ and $\log(1 + \overline{\Gamma}_l)$.

Let $k_0 = [p/2]$. Under the assumption $A_5[2k_0]$, we define the quasi-loglikelihood function of $\mathbf{X}_n$ as

$$l_{p,n}(\theta) := l_{p,n}(\theta, \mathbf{X}_n) := \sum_{i=1}^n \mathtt{l}_{p,i}(\theta) \tag{3.11}$$

where

$$\begin{aligned}
\mathtt{l}_{p,i}(\theta) := &\frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta))^2}{2\Delta_n c_{i-1}(\beta)}\left\{1 + \sum_{j=1}^{k_0} \Delta_n^j \mathrm{d}_j(\theta, X_{i-1})\right\} \\
&+ \frac{1}{2}\left\{\log c_{i-1}(\beta) + \sum_{j=1}^{k_0} \Delta_n^j \mathrm{e}_j(\theta, X_{i-1})\right\}
\end{aligned} \tag{3.12}$$

and $\mathrm{d}_j$, resp. $\mathrm{e}_j$, is the coefficient of $\Delta_n^j$ in the Taylor expansion of $(1+\overline{\Gamma}_{k_0+1}(\Delta_n, x, \theta))^{-1}$, resp. $\log(1 + \overline{\Gamma}_{k_0+1}(\Delta_n, x, \theta))$. It is not hard to show that, for example,

$$\begin{aligned}
\mathrm{d}_1(\theta, x) &= -\,\mathrm{e}_1(\theta, x) = -\frac{\gamma_2(\theta, x)}{c(\beta, x)}, \\
\mathrm{d}_2(\theta, x) &= -\,\mathrm{e}_2(\theta, x) = \frac{1}{c(\beta, x)}\left[\frac{\gamma_2^2(\theta, x)}{c(\beta, x)} - \gamma_3(\theta, x)\right].
\end{aligned}$$

*Remark 3.1* It is worth to point out that by assumptions $A_3$ and $A_5$ emerge that $\mathrm{d}_j$ and $\mathrm{e}_j$, for all $j \leq k_0$, are three times differentiable with respect to $\theta$. Furthermore, all their derivatives with respect to $\theta$ are of polynomial growth in $x$ uniformly in $\theta$.

The contrast function (3.11) yields to the maximum quasi-likelihood estimator $\hat{\theta}_{p,n} := (\hat{\alpha}_{p,n}, \hat{\beta}_{p,n})$ defined as

$$l_{p,n}(\hat{\theta}_{p,n}) = \inf_{\theta \in \Theta} l_{p,n}(\theta). \tag{3.13}$$

Let $I(\theta_0)$ be the Fisher information matrix at $\theta_0$ defined as follows

$$I(\theta_0) := \begin{pmatrix} [I_b^{h,k}(\theta_0)]_{h,k=1,\ldots,m_1} & 0 \\ 0 & [I_\sigma^{h,k}(\theta_0)]_{h,k=1,\ldots,m_2} \end{pmatrix}, \tag{3.14}$$

where

$$I_b^{h,k}(\theta_0) := \int \left(\frac{\partial_{\alpha_h} b\, \partial_{\alpha_k} b}{c}\right)(\theta_0, x)\pi_0(\mathrm{d}x),$$
$$I_\sigma^{h,k}(\theta_0) := \frac{1}{2}\int \left(\frac{\partial_{\beta_h} c\, \partial_{\beta_k} c}{c^2}\right)(\beta_0, x)\pi_0(\mathrm{d}x).$$

We recall an important asymptotic result which will be useful in the proof of our main theorem.

**Theorem 1** (Kessler 1997) *Let $p$ be an integer and $k_0 = [p/2]$. Under assumptions $A_1$ to $A_4$, $A_5[2k_0]$ and $A_6$, if $\Delta_n \to 0, n\Delta_n \to \infty$, as $n \to \infty$, the estimator $\hat{\theta}_{p,n}$ is consistent; i.e.*

$$\hat{\theta}_{p,n} \xrightarrow[n\to\infty]{P_0} \theta_0. \tag{3.15}$$

*If in addition $n\Delta_n^p \to 0$ and $\theta_0 \in Int(\Theta)$ then*

$$\varphi(n)^{-1/2}(\hat{\theta}_{p,n} - \theta_0) = \begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_{p,n} - \alpha_0) \\ \sqrt{n}(\hat{\beta}_{p,n} - \beta_0) \end{pmatrix} \xrightarrow[n\to\infty]{d} N_{m_1+m_2}(0, I^{-1}(\theta_0)), \tag{3.16}$$

*where*

$$\varphi(n) := \begin{pmatrix} \frac{1}{n\Delta_n} I_{m_1} & 0 \\ 0 & \frac{1}{n} I_{m_2} \end{pmatrix}.$$

*Remark 3.2* We observe that $l_{2,n}$ does not coincide with (3.2), because (3.11) contains the terms $\mathrm{d}_1$ and $\mathrm{e}_1$. Nevertheless, $l_n$ also yields an asymptotical efficient estimator for $\theta$ and then we refer to it when $p = 2$.

*Remark 3.3* Under the same framework adopted in this paper, alternatively to $\hat{\theta}_{p,n}$, Kessler (1995) and Uchida and Yoshida (2012) proposed different types of adaptive maximum quasi-likelihood estimators. For instance, in Uchida and Yoshida (2012), the first type of adaptive estimator is introduced starting from the initial estimator $\tilde{\beta}_{0,n}$ given by $\mathbb{U}_n(\tilde{\beta}_{0,n}) = \inf_{\beta\in\Theta_\beta} \mathbb{U}_n(\beta)$, where

$$\mathbb{U}_n(\beta) := \frac{1}{2}\sum_{i=1}^{n} \left\{ \frac{(X_i - X_{i-1})^2}{\Delta_n c_{i-1}(\beta)} + \log c_{i-1}(\beta) \right\}.$$

For $p \geq 2, k_0 = [p/2]$ and $l_0 = [(p-1)/2]$, the first type adaptive estimator $\tilde{\theta}_{p,n} = (\tilde{\alpha}_{k_0,n}, \tilde{\beta}_{l_0,n})$ is defined for $k = 1, 2, \ldots, k_0$, as follows

$$l_{p,n}(\tilde{\alpha}_{k,n}, \tilde{\beta}_{k-1,n}) = \inf_{\alpha\in\Theta_\alpha} l_{p,n}(\alpha, \tilde{\beta}_{k-1,n}),$$
$$l_{p,n}(\tilde{\alpha}_{k,n}, \tilde{\beta}_{k,n}) = \inf_{\beta\in\Theta_\beta} l_{p,n}(\tilde{\alpha}_{k,n}, \beta).$$

The maximum quasi-likelihood estimator $\hat{\theta}_{p,n}$ and its adaptive versions, like $\tilde{\theta}_{p,n}$, are asymptotically equivalent (under a minor change of the initial assumptions); i.e. they have the same properties (3.15) and (3.16) (see Uchida and Yoshida 2012). In what follow we will developed a test based on $\hat{\theta}_{p,n}$; nevertheless in light of the previous discussion, it would be possible to replace $\hat{\theta}_{p,n}$ with $\tilde{\theta}_{p,n}$.

## 4 Test statistics

The goal of this section is to introduce a new type of test statistics for the following parametric hypotheses problem

$$H_0 : \theta = \theta_0, \quad \text{vs} \quad H_1 : \theta \neq \theta_0, \tag{4.1}$$

concerning the stochastic differential equation (1.1). $X$ is partially observed and therefore we have discrete observations represented by $\mathbf{X}_n$. The motivation of this research is due to the fact that under non-simple alternative hypotheses do not exist uniformly most powerful parametric tests. Therefore, we need proper procedure for making the right decision concerning statistical hypothesis.

The first step consists in the introduction of a suitable measure regarding the "discrepancy", or the "distance", between diffusions belonging to the parametric class (1.1). Furthermore, as recalled in the previous section, for a general stochastic differential equation $X$, the true probability transitions from $X_{i-1}$ to $X_i$ do not exist in closed form as well as the likelihood function. Suppose known the parameter $\beta$ and assume observable the sample path up to time $T = n\Delta_n$. Let $Q_\beta$ be the probability law of the process solution to $\mathrm{d}Y_t = \sigma(\beta, Y_t)\mathrm{d}W_t$. The continuous log-likelihood of $X$ is given by

$$\log \frac{\mathrm{d}P_\theta}{\mathrm{d}Q_\beta} = \int_0^T \frac{b(\alpha, X_t)}{c(\beta, X_t)}\mathrm{d}X_t - \frac{1}{2}\int_0^T \frac{b^2(\alpha, X_t)}{c(\beta, X_t)}\mathrm{d}t.$$

Thus we can consider the (squared) $L^2(Q_\beta)$-distance between the log-likelihoods $\log \frac{\mathrm{d}P_{\theta_1}}{\mathrm{d}Q_\beta}$ and $\log \frac{\mathrm{d}P_{\theta_2}}{\mathrm{d}Q_\beta}$ with $\theta_1, \theta_2 \in \Theta$; that is

$$D(\theta_1, \theta_2) := \left\| \log \frac{\mathrm{d}P_{\theta_1}}{\mathrm{d}Q_\beta} - \log \frac{\mathrm{d}P_{\theta_2}}{\mathrm{d}Q_\beta} \right\|^2_{L^2(Q_\beta)} = \int \left[ \log \frac{\mathrm{d}P_{\theta_1}}{\mathrm{d}Q_\beta} - \log \frac{\mathrm{d}P_{\theta_2}}{\mathrm{d}Q_\beta} \right]^2 \mathrm{d}Q_\beta. \tag{4.2}$$

Clearly for testing the hypotheses (4.1) in the framework of discretely observed stochastic differential equations, the distance (4.2) is not useful. Nevertheless, the above $L^2-$metric for the continuos observations suggests to consider

$$\mathbb{D}_{p,n}(\theta_1, \theta_2) := \frac{1}{n}\sum_{i=1}^n [\mathbb{l}_{p,i}(\theta_1) - \mathbb{l}_{p,i}(\theta_2)]^2, \quad \theta_1, \theta_2 \in \Theta, \tag{4.3}$$

which can be interpreted as the empirical version of (4.2), where the theoretical log-likelihood is replaced with the quasi-log-likelihood defined by (3.11). The following theorem provides the convergence in probability of $\mathbb{D}_{p,n}$.

**Theorem 2** *Let $p$ be an integer and $k_0 = [p/2]$. Assume $A_1 - A_4$, $A_5[2k_0]$ and $A_6$. Under $H_0$, if $\Delta_n \to 0, n\Delta_n \to \infty$, as $n \to \infty$, we have that*

$$\mathbb{D}_{p,n}(\theta, \theta_0) \xrightarrow[n\to\infty]{P_0} U(\beta, \beta_0)$$

*uniformly in* $\theta$, *where*

$$U(\beta,\beta_0):=\frac{1}{4}\int\left\{3\left[\frac{c(\beta_0,x)}{c(\beta,x)}-1\right]^2+\left[\log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)\right]^2\right.$$
$$\left.+\left[\frac{c(\beta_0,x)}{c(\beta,x)}-1\right]\log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)\right\}\pi_0(\mathrm{d}x).$$

The above result shows that $\mathbb{D}_{p,n}(\theta,\theta_0)$ is not a true approximation of $D_{p,n}(\theta,\theta_0)$ because it does not converge to $\int\left[\log(\pi_\theta(\mathrm{d}x)/\pi_0(\mathrm{d}x))\right]^2\pi_0(\mathrm{d}x)$. Nevertheless, the function (4.3) allows to construct the main object of interest of the paper. Let $\hat{\theta}_n$ be the maximum quasi-likelihood estimator defined by (3.13), for testing the hypotheses (4.1) we introduce the following class of test statistics

$$T_{p,n}(\hat{\theta}_{p,n},\theta_0):=n\mathbb{D}_{p,n}(\hat{\theta}_{p,n},\theta_0). \tag{4.4}$$

The first result concerns the weak convergence of $T_{p,n}(\hat{\theta}_{p,n},\theta_0)$. We prove that $T_{p,n}(\hat{\theta}_{p,n},\theta_0)$ is asymptotically distribution free under $H_0$; namely it weakly converges to a chi-squared random variable with $m_1+m_2$ degrees of freedom.

**Theorem 3** *Let* $p$ *be an integer and* $k_0=[p/2]$. *Assume* $A_1-A_4$, $A_5[2k_0]$ *and* $A_6$. *Under* $H_0$, *if* $\Delta_n\to 0$, $n\Delta_n\to\infty$, $n\Delta_n^p\to 0$, *as* $n\to\infty$, *we have that*

$$T_{p,n}(\hat{\theta}_{p,n},\theta_0)\underset{n\to\infty}{\overset{d}{\longrightarrow}}\chi^2_{m_1+m_2}. \tag{4.5}$$

Given the level $\alpha\in(0,1)$, our criterion suggests to

$$\text{reject } H_0 \text{ if } \; T_{p,n}(\hat{\theta}_{p,n},\theta_0)>\chi^2_{m_1+m_2,\alpha},$$

where $\chi^2_{m_1+m_2,\alpha}$ is the $1-\alpha$ quantile of the limiting random variable $\chi^2_{m_1+m_2}$; that is under $H_0$

$$\lim_{n\to\infty}P_\theta(T_{p,n}(\hat{\theta}_{p,n},\theta_0)>\chi^2_{m_1+m_2,\alpha})=\alpha.$$

Under $H_1$, the power function of the proposed test are equal to the following map

$$\theta\mapsto P_\theta\left(T_{p,n}(\hat{\theta}_{p,n},\theta_0)>\chi^2_{m_1+m_2,\alpha}\right)$$

Often a way to judge the quality of sequences of tests is provided by the powers at alternatives that become closer and closer to the null hypothesis. This justify the study of local limiting power. Indeed, usually the power functions of test statistic (4.4) cannot be calculated explicitly. Nevertheless, $P_\theta\left(T_{p,n}(\hat{\theta}_{p,n},\theta_0)>\chi^2_{m_1+m_2,\alpha}\right)$ can be studied and approximated under contiguous alternatives written as

$$H_{1,n}:\theta=\theta_0+\varphi(n)^{1/2}h, \tag{4.6}$$

where $h\in\mathbb{R}^{m_1+m_2}$ such that $\theta_0+\varphi(n)^{1/2}h\in\Theta$. In order to get a reasonable approximation of the power function, we analyze the asymptotic law of the test statistics under the local alternatives $H_{1,n}$. We need the following assumption on the contiguity of probability measures (see Van der Vaart 1998):

$B_1$ $P_{\theta_0+\varphi(n)^{1/2}h}$ is a sequence of contiguous probability measures with respect to $P_0$; i.e. $\lim_{n\to\infty}P_0(A_n)=0$ implies $\lim_{n\to\infty}P_{\theta_0+\varphi(n)^{1/2}h}(A_n)=0$ for every measurable sets $A_n$.

*Remark 4.1* The assumption $B_1$ holds if we assume $A_1 - A_4$, $A_5[2k_0]$ and the conditions:

(i) there exists a constant $C > 0$ such that the following estimates hold

$$|b(\alpha, x)| \leq C(1 + |x|), \quad \left|\frac{\partial}{\partial x} b(\alpha, x)\right| + |\sigma(\beta, x)| + \left|\frac{\partial}{\partial x}\sigma(\beta, x)\right| \leq C$$

for all $(\alpha, \beta) \in \Theta$ and $x \in \mathbb{R}$;

(ii) there exists $C_0 > 0$ and $K > 0$ such that

$$b(\alpha, x)x \leq -C_0|x|^2 + K$$

for all $(\alpha, x) \in \Theta_\alpha \times \mathbb{R}$;

(iii) there exists a constant $C_1 > 1$ such that

$$\frac{1}{C_1} \leq \sigma(\beta, x) \leq C_1, \text{ for all } (\beta, x) \in \Theta_\beta \times \mathbb{R}.$$

Under the above assumptions, Gobet (2002) proved the Local Asymptotic Normality (LAN) for the likelihood of the ergodic diffusions (1.1); i.e.

$$\log\left(\frac{\mathrm{d}P_{\theta_0+\varphi(n)^{1/2}h}}{\mathrm{d}P_0}(\mathbf{X}_n)\right) \xrightarrow[n\to\infty]{d} h' N_{m_1+m_2}(0, I(\theta_0)) + \frac{1}{2}h' I(\theta_0)h.$$

By means of Le Cam's first lemma (see Van der Vaart 1998), LAN property implies the contiguity of $P_{\theta_0+\varphi(n)^{1/2}h}$ with respect to $P_0$.

Now, we are able to study the asymptotic probability distribution of $T_{p,n}$ under $H_{1,n}$.

**Theorem 4** *Let $p$ be an integer and $k_0 = [p/2]$. Assume $A_1 - A_4$, $A_5[2k_0]$, $A_6$ and $B_1$ fulfill. Under the local alternative hypothesis $H_{1,n}$, if $\Delta_n \to 0, n\Delta_n \to \infty, n\Delta_n^p \to 0$ as $n \to \infty$, the following weak convergence holds*

$$T_{p,n}(\hat{\theta}_{p,n}, \theta_0) \xrightarrow[n\to\infty]{d} \chi^2_{m_1+m_2}(h' I(\theta_0)h), \tag{4.7}$$

*where the random variable $\chi^2_{l+m}(h' I(\theta_0)h)$ is a non-central chi square random variable with $l + m$ degrees of freedom and non-centrality parameter $h' I(\theta_0)h$.*

*Remark 4.2* If we deal with $H_0 : \theta = \theta_0$ and the local alternative hypothesis $H_{1,n}$, Theorem 4 leads to the following approximation of the power functions

$$P_\theta\left(T_{p,n}(\hat{\theta}_{p,n}, \theta_0) > \chi^2_{m_1+m_2,\alpha}\right) \cong 1 - \mathbf{F}\left(\chi^2_{m_1+m_2,\alpha}\right), \quad n >> 1, \tag{4.8}$$

where $\mathbf{F}(\cdot)$ is the cumulative function of the random variable $\chi^2_{m_1+m_2}(h' I(\theta_0)h)$.

*Remark 4.3* The Generalized Quasi-Likelihood Ratio, Wald, Rao type test statistics have been studied by Kitagawa and Uchida (2014). These test statistics are, respectively, defined as follows

$$L_{p,n}(\hat{\theta}_{p,n}, \theta_0) := 2(l_{p,n}(\hat{\theta}_{p,n}) - l_{p,n}(\theta_0)) \tag{4.9}$$

$$W_{p,n}(\hat{\theta}_{p,n}, \theta_0) := (\varphi(n)^{-1/2}(\hat{\theta}_{p,n} - \theta_0))' I_{p,n}(\hat{\theta}_{p,n})\varphi(n)^{-1/2}(\hat{\theta}_{p,n} - \theta_0) \tag{4.10}$$

$$R_{p,n}(\hat{\theta}_{p,n}, \theta_0) := (\varphi(n)^{1/2}\partial_\theta l_{p,n}(\theta_0))' I^{-1}_{p,n}(\hat{\theta}_{p,n})\varphi(n)^{1/2}\partial_\theta l_{p,n}(\theta_0), \tag{4.11}$$

where

$$I_{p,n}(\theta) = \begin{pmatrix} \frac{1}{n\Delta_n}\partial^2_\alpha l_{p,n}(\theta) & \frac{1}{n\sqrt{\Delta_n}}\partial_\alpha\partial_\beta l_{p,n}(\theta) \\ \frac{1}{n\sqrt{\Delta_n}}\partial_\beta\partial_\alpha l_{p,n}(\theta) & \frac{1}{n}\partial^2_\beta l_{p,n}(\theta) \end{pmatrix}$$

and $R_{p,n}$ is well-defined if $I_{p,n}(\theta)$ is nonsingular. The above test statistics are asymptotically equivalent to $T_{p,n}$; i.e. under $H_0$, $L_{p,n}$, $W_{p,n}$ and $R_{p,n}$ weakly converge to a $\chi^2$ random variable.

*Remark 4.4* In De Gregorio and Iacus (2013), the authors dealt with (for $p = 2$) test statistics based on an empirical version of the true $\phi$-divergences; i.e.

$$2\sum_{i=1}^{n}\phi\left(\frac{\exp l_n(\theta)}{\exp l_n(\theta_0)}\right) \tag{4.12}$$

where $\phi$ represents a suitable convex function and $l_n$ is given by (3.2). In the present paper, the starting point is represented by the $L^2$-distance between two diffusion parametric models. Somehow, the approach developed in this work is close to that developed by Aït-Sahalia (1996), where a test based on the $L^2$-distance measure between the density function and its nonparametric estimator is introduced.

*Remark 4.5* From a practical point of view, since sometimes $\alpha = \alpha_0$ and $\beta = \beta_0$ have different meanings, it is possible to deal with a stepwise procedure. For instance as $p = 2$, first, we test $\beta = \beta_0$ by means of

$$T_n^{\beta}(\tilde{\beta}_{0,n}, \beta_0) := \sum_{i=1}^{n}\left[\frac{(X_i - X_{i-1})^2}{\Delta_n}\left(\frac{1}{c_{i-1}(\tilde{\beta}_{0,n})} - \frac{1}{c_{i-1}(\beta_0)}\right) + \log\left(\frac{c_{i-1}(\tilde{\beta}_{0,n})}{c_{i-1}(\beta_0)}\right)\right]^2$$

and then, in the second step, we test $\alpha = \alpha_0$ by taking into account

$$T_n^{\alpha}(\tilde{\alpha}_{1,n}, \alpha_0, \tilde{\beta}_{0,n}) := \sum_{i=1}^{n}[\mathbb{1}_{2,i}(\tilde{\alpha}_{1,n}, \tilde{\beta}_{0,n}) - \mathbb{1}_{2,i}(\alpha_0, \tilde{\beta}_{0,n})]^2,$$

where $\tilde{\alpha}_{1,n}$ and $\tilde{\beta}_{0,n}$ are the adaptive estimators defined in the Remark 3.3.

## 5 Numerical analysis

Although all test statistics presented in the above and in the literature satisfy the same asymptotic results, for small sample sizes the performance of each test statistic is determined by the statistical model generating the data and the quality of the approximation of the quasi-likelihood function. To put in evidence these effects we consider the two stochastic models presented in Sect. 2, namely the Ornstein-Uhlenbeck (OU in the tables) of Eq. (2.1) and the CIR model of Eq. (2.2). In this numerical study we consider the power of the test under local alternatives for different test statistics:

- the $\phi$ divergence of Eq. (4.12) with $\phi(x) = 1 - x + x\log(x)$, which is equivalent to the approximated Kullback–Leibler divergence (see, De Gregorio and Iacus 2013). We use the label $AKL$ in the tables for this approximate Kullback-Leibler measure;
- the $\phi$ divergence with $\phi(x) = \left(\frac{x-1}{x+1}\right)^2$: this was proposed in Balakrishnan and Sanghvi (1968), we name it BS in the tables;
- the Generalized Quasi-Likelihood Ratio test with $p = 2$, see e.g., (4.9), denoted as GQLRT in the tables;
- the Rao test statistics[1] $R(\hat{\theta}_{p,n}, \theta_0)$ of Eq. (4.11), denoted as RAO in the tables;

[1] We do not consider the Wald test of (4.10) because it was shown in Kitagawa and Uchida (2014) that it performs similarly to the Rao test statistics.

- and the statistic $T_{p,n}(\hat{\theta}_{p,n}, \theta_0)$ proposed in this paper and defined in Eq. (4.4), with $p = 2$, denoted as $T_{2,n}$ in the tables.

The sample sizes have been chosen to be equal to $n = 50, 100, 250, 500, 1000$ observations and time horizon is set to $T = n^{\frac{1}{3}}$, in order to satisfy the asymptotic theory. For testing $\theta_0$ against the local alternatives $\theta_0 + \frac{h}{\sqrt{n\Delta_n}}$ for the parameters in the drift coefficient and $\theta_0 + \frac{h}{\sqrt{n}}$ for the parameters in the diffusion coefficient, $h$ is taken in a grid from 0 to 1, and $h = 0$ corresponds to the null hypothesis $H_0$. For the data generating process, we consider the following statistical models

OU the one-dimensional Ornstein–Uhlenbeck model solution to $\mathrm{d}X_t = \alpha_1(\alpha_2 - X_t)\mathrm{d}t + \beta_1 \mathrm{d}W_t$, $X_0 = 1$, with $\theta_0 = (\alpha_1, \alpha_2, \beta_1) = (0.5, 0.5, 0.25)$;

CIR the one-dimensional CIR model solution to $\mathrm{d}X_t = \alpha_1(\alpha_2 - X_t)\mathrm{d}t + \beta_1\sqrt{X_t}\mathrm{d}W_t$, $X_0 = 1$, with $\theta_0 = (\alpha_1, \alpha_2, \beta_1) = (0.5, 0.5, 0.125)$.

In each experiments the process have been simulated at high frequency using the Euler-Maruyama scheme and resampled to obtain $n = 50, 100, 250, 500, 1000$ observations. Remark that, even if the Ornstein-Uhlenbeck process has a Gaussian transition density, this density is different from the Euler-Maruyama Gaussian density for non negligible time mesh $\Delta_n$ (see, Iacus 2008). For the simulation we used the R package *yuima* (see, Iacus and Yoshida 2017). Each experiment is replicated 1000 times and from the empirical distribution of each test statistic, say $S_n$, we define the rejection threshold of the test as $\tilde{\chi}^2_{3,0.05}$, i.e. $\tilde{\chi}^2_{3,0.05}$ is the 95% quantile of the empirical distribution of $S_n$, that is

$$0.05 = \mathrm{Freq}\left(S_n(\hat{\theta}_n, \theta_0) > \tilde{\chi}^2_{3,0.05}\right).$$

Similarly, we define the empirical power function of the test as

$$\mathrm{EPow}(h) = \mathrm{Freq}\left(S_n(\hat{\theta}_n, \theta_0 + \varphi(n)^{1/2}h) > \tilde{\chi}^2_{3,0.05}\right),$$

where $\hat{\theta}_n$ is the maximum quasi-likelihood estimator defined in (3.13). The choice of using the empirical threshold $\tilde{\chi}^2_{3,0.05}$ instead of the theoretical threshold $\chi^2_{3,0.05}$ from the $\chi^2_3$ distribution, is due to the fact that otherwise the tests are non comparable. Indeed, the empirical level of the test is not 0.05 for small sample sizes when $\chi^2_{3,0.05}$ is used as rejection threshold and, for example, when $h = 0$ different choices of the test statistic produce different empirical levels of the test. Tables 1 and 2 contain the empirical power function of each test. In these tables the bold face font is used to put in evidence the test statistics with the highest empirical power function $\mathrm{EPow}(h)$ for a given local alternative $h > 0$.

From this numerical analysis we can see several facts:

- the test statistic based on the AKL does not perform as the GQLR test despite they are related to the same divergence; the latter being sometimes better;
- the $T_{2,n}$ seems to be (almost) uniformly more powerful in this experiment;
- all but RAO test seem to have a good behaviour when the alternative is sufficiently large;
- for the CIR model, the RAO test does not perform well under the alternative hypothesis and this is probably because it requires very large $T$ which, in our case, is at most $T = 10$. For the OU Gaussian case, the performance are better and in line from those presented in Kitagawa and Uchida (2014) for similar sample sizes.

Therefore, we can conclude that, despite all the test statistics share the same asymptotic properties, the proposed $T_{p,n}$ seems to perform very well in the small sample case examined in the above Monte Carlo experiments, at least for $p = 2$.

**Table 1** Empirical power function EPow($h$), for different sample sizes $n$ and local alternatives $h$

| | AKL | GQLRT | BS | RAO | $T_{2,n}$ |
|---|---|---|---|---|---|
| $n = 50$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.044 | 0.048 | 0.046 | **0.053** | 0.052 |
| h = 0.05 | 0.035 | 0.032 | 0.041 | **0.057** | **0.057** |
| h = 0.10 | 0.025 | 0.029 | 0.033 | 0.064 | **0.077** |
| h = 0.20 | 0.011 | 0.031 | 0.042 | 0.078 | **0.133** |
| h = 0.30 | 0.007 | 0.054 | 0.069 | 0.096 | **0.239** |
| h = 0.40 | 0.007 | 0.108 | 0.147 | 0.121 | **0.371** |
| h = 0.50 | 0.009 | 0.216 | 0.269 | 0.138 | **0.559** |
| h = 0.60 | 0.021 | 0.359 | 0.448 | 0.146 | **0.720** |
| h = 0.70 | 0.053 | 0.527 | 0.591 | 0.149 | **0.842** |
| h = 0.80 | 0.120 | 0.670 | 0.736 | 0.150 | **0.917** |
| h = 0.90 | 0.221 | 0.794 | 0.852 | 0.148 | **0.966** |
| h = 1.00 | 0.383 | 0.882 | 0.910 | 0.145 | **0.992** |
| $n = 100$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.046 | 0.047 | 0.046 | **0.050** | **0.050** |
| h = 0.05 | 0.032 | 0.035 | 0.035 | 0.050 | **0.055** |
| h = 0.10 | 0.022 | 0.029 | 0.030 | 0.058 | **0.070** |
| h = 0.20 | 0.014 | 0.038 | 0.042 | 0.082 | **0.141** |
| h = 0.30 | 0.009 | 0.089 | 0.083 | 0.101 | **0.253** |
| h = 0.40 | 0.009 | 0.159 | 0.163 | 0.128 | **0.404** |
| h = 0.50 | 0.020 | 0.283 | 0.291 | 0.155 | **0.609** |
| h = 0.60 | 0.051 | 0.465 | 0.472 | 0.183 | **0.769** |
| h = 0.70 | 0.131 | 0.644 | 0.659 | 0.199 | **0.876** |
| h = 0.80 | 0.244 | 0.789 | 0.801 | 0.213 | **0.943** |
| h = 0.90 | 0.414 | 0.883 | 0.893 | 0.221 | **0.984** |
| h = 1.00 | 0.608 | 0.937 | 0.944 | 0.225 | **0.996** |
| $n = 250$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.044 | 0.049 | 0.050 | **0.051** | 0.048 |
| h = 0.05 | 0.036 | 0.049 | 0.046 | 0.052 | **0.057** |
| h = 0.10 | 0.028 | 0.048 | 0.050 | 0.058 | **0.075** |
| h = 0.20 | 0.015 | 0.076 | 0.078 | 0.114 | **0.143** |
| h = 0.30 | 0.022 | 0.153 | 0.157 | 0.168 | **0.255** |
| h = 0.40 | 0.049 | 0.304 | 0.304 | 0.222 | **0.452** |
| h = 0.50 | 0.118 | 0.486 | 0.496 | 0.280 | **0.654** |
| h = 0.60 | 0.253 | 0.703 | 0.704 | 0.339 | **0.822** |
| h = 0.70 | 0.436 | 0.847 | 0.851 | 0.389 | **0.921** |
| h = 0.80 | 0.666 | 0.928 | 0.931 | 0.419 | **0.969** |
| h = 0.90 | 0.821 | 0.973 | 0.976 | 0.462 | **0.991** |
| h = 1.00 | 0.911 | 0.992 | 0.993 | 0.485 | **1.000** |

**Table 1** continued

| | AKL | GQLRT | BS | RAO | $T_{2,n}$ |
|---|---|---|---|---|---|
| $n = 500$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.048 | 0.049 | 0.049 | **0.052** | 0.051 |
| h = 0.05 | 0.038 | 0.044 | 0.043 | **0.067** | 0.059 |
| h = 0.10 | 0.032 | 0.050 | 0.050 | **0.082** | 0.075 |
| h = 0.20 | 0.030 | 0.084 | 0.080 | **0.134** | 0.133 |
| h = 0.30 | 0.050 | 0.175 | 0.175 | 0.202 | **0.250** |
| h = 0.40 | 0.138 | 0.329 | 0.323 | 0.279 | **0.449** |
| h = 0.50 | 0.274 | 0.555 | 0.552 | 0.363 | **0.673** |
| h = 0.60 | 0.493 | 0.751 | 0.747 | 0.454 | **0.828** |
| h = 0.70 | 0.704 | 0.869 | 0.869 | 0.522 | **0.934** |
| h = 0.80 | 0.847 | 0.957 | 0.957 | 0.584 | **0.983** |
| h = 0.90 | 0.936 | 0.987 | 0.987 | 0.630 | **0.996** |
| h = 1.00 | 0.982 | 0.997 | 0.997 | 0.678 | **0.998** |
| $n = 1000$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.046 | 0.049 | 0.050 | **0.051** | **0.051** |
| h = 0.05 | 0.038 | 0.046 | 0.049 | 0.056 | **0.058** |
| h = 0.10 | 0.035 | 0.056 | 0.062 | 0.062 | **0.074** |
| h = 0.20 | 0.061 | 0.104 | 0.109 | 0.121 | **0.134** |
| h = 0.30 | 0.122 | 0.182 | 0.187 | 0.193 | **0.241** |
| h = 0.40 | 0.219 | 0.359 | 0.372 | 0.291 | **0.442** |
| h = 0.50 | 0.426 | 0.600 | 0.605 | 0.398 | **0.662** |
| h = 0.60 | 0.655 | 0.786 | 0.794 | 0.507 | **0.840** |
| h = 0.70 | 0.821 | 0.912 | 0.914 | 0.596 | **0.942** |
| h = 0.80 | 0.930 | 0.969 | 0.972 | 0.665 | **0.985** |
| h = 0.90 | 0.978 | 0.993 | 0.993 | 0.711 | **0.994** |
| h = 1.00 | 0.994 | 0.997 | 0.997 | 0.760 | **0.998** |

The empirical power and theoretical power is 0.05. Data generating model: the 1-dimensional Ornstein–Uhlenbeck process

# 6 Proofs

In order to prove the theorems appearing in the paper, we need some preliminary results. Let us start with the following lemmas.

**Lemma 1** *For $k \geq 1$ and $t_{i-1}^n \leq t \leq t_i^n$*

$$E_0^{i-1}[|X_t - X_{i-1}|^k] \leq C_k |t - t_{i-1}^n|^{k/2}(1 + |X_{i-1}|)^{C_k}. \tag{6.1}$$

*If $f : \Theta \times \mathbb{R} \to \mathbb{R}$ is of polynomial growth in $x$ uniformly in $\theta$ then*

$$E_0^{i-1}[f(\theta, X_t)] \leq C_{t-t_{i-1}^n}(1 + |X_{i-1}|)^C, \quad t_{i-1}^n \leq t \leq t_i^n. \tag{6.2}$$

*Proof* See the proof of Lemma 6 in Kessler (1997). □

**Table 2** Empirical power function EPow($h$), for different sample sizes $n$ and local alternatives $h$

| | AKL | GQLRT | BS | RAO | $T_{2,n}$ |
|---|---|---|---|---|---|
| $n = 50$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.041 | 0.044 | 0.045 | 0.052 | **0.053** |
| h = 0.05 | 0.025 | 0.032 | 0.031 | 0.059 | 0.**071** |
| h = 0.10 | 0.009 | 0.040 | 0.042 | 0.068 | **0.145** |
| h = 0.20 | 0.013 | 0.148 | 0.167 | 0.075 | **0.371** |
| h = 0.30 | 0.044 | 0.416 | 0.458 | 0.069 | **0.721** |
| h = 0.40 | 0.186 | 0.700 | 0.741 | 0.067 | **0.923** |
| h = 0.50 | 0.475 | 0.883 | 0.907 | 0.067 | **0.989** |
| h = 0.60 | 0.760 | 0.967 | 0.981 | 0.061 | **0.997** |
| h = 0.70 | 0.913 | 0.994 | 0.998 | 0.059 | **1.000** |
| h = 0.80 | 0.981 | **1.000** | **1.000** | 0.051 | **1.000** |
| h = 0.90 | 0.997 | **1.000** | **1.000** | 0.041 | **1.000** |
| h = 1.00 | **1.000** | **1.000** | **1.000** | 0.041 | **1.000** |
| $n = 100$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.040 | 0.043 | 0.046 | **0.053** | **0.051** |
| h = 0.05 | 0.019 | 0.032 | 0.034 | 0.056 | **0.070** |
| h = 0.10 | 0.010 | 0.054 | 0.051 | 0.062 | **0.150** |
| h = 0.20 | 0.017 | 0.205 | 0.207 | 0.063 | **0.461** |
| h = 0.30 | 0.102 | 0.537 | 0.553 | 0.064 | **0.797** |
| h = 0.40 | 0.338 | 0.827 | 0.836 | 0.064 | **0.957** |
| h = 0.50 | 0.685 | 0.950 | 0.958 | 0.063 | **0.995** |
| h = 0.60 | 0.896 | 0.993 | 0.994 | 0.059 | **1.000** |
| h = 0.70 | 0.977 | 0.999 | 0.998 | 0.056 | **1.000** |
| h = 0.80 | 0.998 | **1.000** | **1.000** | 0.053 | **1.000** |
| h = 0.90 | 0.999 | **1.000** | **1.000** | 0.048 | **1.000** |
| h = 1.00 | **1.000** | **1.000** | **1.000** | 0.044 | 1.000 |
| $n = 250$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.042 | 0.049 | 0.046 | **0.052** | 0.050 |
| h = 0.05 | 0.026 | 0.045 | 0.046 | 0.054 | **0.071** |
| h = 0.10 | 0.021 | 0.086 | 0.084 | 0.057 | **0.144** |
| h = 0.20 | 0.093 | 0.347 | 0.342 | 0.062 | **0.505** |
| h = 0.30 | 0.372 | 0.752 | 0.756 | 0.064 | **0.864** |
| h = 0.40 | 0.790 | 0.943 | 0.944 | 0.065 | **0.977** |
| h = 0.50 | 0.952 | 0.994 | 0.994 | 0.064 | **1.000** |
| h = 0.60 | 0.996 | **1.000** | **1.000** | 0.060 | **1.000** |
| h = 0.70 | **1.000** | **1.000** | **1.000** | 0.060 | **1.000** |
| h = 0.80 | **1.000** | **1.000** | **1.000** | 0.057 | **1.000** |

**Table 2** continued

| | AKL | GQLRT | BS | RAO | $T_{2,n}$ |
|---|---|---|---|---|---|
| h = 0.90 | **1.000** | **1.000** | **1.000** | 0.055 | **1.000** |
| h = 1.00 | **1.000** | **1.000** | **1.000** | 0.050 | **1.000** |
| $n = 500$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.043 | 0.043 | 0.042 | **0.051** | 0.048 |
| h = 0.05 | 0.030 | 0.046 | 0.044 | 0.051 | **0.074** |
| h = 0.10 | 0.032 | 0.095 | 0.091 | 0.052 | **0.147** |
| h = 0.20 | 0.180 | 0.384 | 0.380 | 0.055 | **0.530** |
| h = 0.30 | 0.598 | 0.802 | 0.800 | 0.058 | **0.869** |
| h = 0.40 | 0.898 | 0.972 | 0.972 | 0.058 | **0.990** |
| h = 0.50 | 0.992 | **0.998** | **0.998** | 0.059 | **0.998** |
| h = 0.60 | 0.998 | **0.999** | **0.999** | 0.057 | **0.999** |
| h = 0.70 | 0.999 | **1.000** | **1.000** | 0.056 | **1.000** |
| h = 0.80 | **1.000** | **1.000** | **1.000** | 0.055 | **1.000** |
| h = 0.90 | **1.000** | **1.000** | **1.000** | 0.055 | **1.000** |
| h = 1.00 | **1.000** | **1.000** | **1.000** | 0.051 | **1.000** |
| $n = 1000$ | | | | | |
| h = 0.00 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| h = 0.01 | 0.044 | 0.048 | 0.047 | **0.051** | 0.050 |
| h = 0.05 | 0.035 | 0.059 | 0.057 | 0.051 | **0.079** |
| h = 0.10 | 0.067 | 0.120 | 0.118 | 0.054 | **0.144** |
| h = 0.20 | 0.274 | 0.429 | 0.428 | 0.058 | **0.527** |
| h = 0.30 | 0.725 | 0.844 | 0.840 | 0.061 | **0.886** |
| h = 0.40 | 0.953 | 0.983 | 0.983 | 0.062 | **0.989** |
| h = 0.50 | 0.996 | 0.998 | 0.998 | 0.062 | **0.999** |
| h = 0.60 | **1.000** | **1.000** | **1.000** | 0.062 | **1.000** |
| h = 0.70 | **1.000** | **1.000** | **1.000** | 0.060 | **1.000** |
| h = 0.80 | **1.000** | **1.000** | **1.000** | 0.059 | **1.000** |
| h = 0.90 | **1.000** | **1.000** | **1.000** | 0.059 | **1.000** |
| h = 1.00 | **1.000** | **1.000** | **1.000** | 0.058 | **1.000** |

The empirical power and theoretical power is 0.05. Data generating model: the 1-dimensional CIR process

**Lemma 2** *For* $l \geq 1$

$$r_l(\Delta_n, X_{i-1}, \theta) = X_{i-1} + \Delta_n b_{i-1}(\alpha) + R(\theta, \Delta_n^2, X_{i-1}) \tag{6.3}$$

$$E_0^{i-1}[(X_i - r_l(\Delta_n, X_{i-1}, \theta))^2] = \Delta_n c_{i-1}(\beta_0) + R(\theta, \Delta_n^2, X_{i-1}) \tag{6.4}$$

$$E_0^{i-1}[(X_i - r_l(\Delta_n, X_{i-1}, \theta))^3] = R(\theta, \Delta_n^2, X_{i-1}) \tag{6.5}$$

$$E_0^{i-1}[(X_i - r_l(\Delta_n, X_{i-1}, \theta))^4] = 3\Delta_n^2 c_{i-1}^2(\beta_0) + R(\theta, \Delta_n^3, X_{i-1}) \tag{6.6}$$

$$E_0^{i-1}[(X_i - r_l(\Delta_n, X_{i-1}, \theta))^5] = R(\theta, \Delta_n^3, X_{i-1}) \tag{6.7}$$

$$E_0^{i-1}[(X_i - r_l(\Delta_n, X_{i-1}, \theta))^6] = 5 \cdot 3\Delta_n^3 c_{i-1}^3(\beta_0) + R(\theta, \Delta_n^4, X_{i-1}) \tag{6.8}$$

$$E_0^{i-1}[(X_i - r_l(\Delta_n, X_{i-1}, \theta))^7] = R(\theta, \Delta_n^4, X_{i-1}) \tag{6.9}$$

$$E_0^{i-1}[(X_i - r_l(\Delta_n, X_{i-1}, \theta))^8] = 7 \cdot 5 \cdot 3\Delta_n^4 c_{i-1}^4(\beta_0) + R(\theta, \Delta_n^5, X_{i-1}) \tag{6.10}$$

*Proof* The equalities from (6.3) to (6.6) represent the statement of Lemma 7 in Kessler (1997). By using the same approach adopted for the proof of the aforementioned lemma, we observe that from (6.3) to (6.6), the result (6.7) and (6.8) hold, if we are able to show that

$$E_0^{i-1}[(X_i - X_{i-1})^5] = R(\theta, \Delta_n^3, X_{i-1}) \tag{6.11}$$

$$E_0^{i-1}[(X_i - X_{i-1})^6] = 5 \cdot 3\Delta_n^3 c_{i-1}^3(\beta_0) + R(\theta, \Delta_n^4, X_{i-1}) \tag{6.12}$$

We only prove (6.12), because (6.11) follows by means of similar arguments. By applying the Itô-Taylor formula [see Lemma 1, in Florens-Zmirou (1989)] to the function $f_x(y) = (y-x)^6$ we obtain

$$\begin{aligned} E_0^{i-1}\big[(X_i - X_{i-1})^6\big] &= f_{X_{i-1}}(X_{i-1}) + \Delta_n L_0 f_{X_{i-1}}(X_{i-1}) \\ &\quad + \frac{\Delta_n^2}{2} L_0^2 f_{X_{i-1}}(X_{i-1}) + \frac{\Delta_n^3}{3!} L_0^3 f_{X_{i-1}}(X_{i-1}) \\ &\quad + \int_0^{\Delta_n}\int_0^{u_1}\int_0^{u_2}\int_0^{u_3} E_0^{i-1}\big[L_0^4 f_{X_{i-1}}(X_{t_{i-1}^n+u_4})\big]\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_3\mathrm{d}u_4. \end{aligned}$$

By applying (6.2), we obtain

$$\int_0^{\Delta_n}\int_0^{u_1}\int_0^{u_2}\int_0^{u_3} E_0^{i-1}\big[L_0^4 f_{X_{i-1}}(X_{t_{i-1}^n+u_4})\big]\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_3\mathrm{d}u_4 = R(\theta, \Delta_n^4, X_{i-1}).$$

Furthermore, by means of long and cumbersome calculations, we can show that $f_x(x) = L_0 f_x(x) = L_0^2 f_x(x) = 0$, while $L_0^3 f_x(x) = 5 \cdot 3 \cdot 3! c_{i-1}^3(\beta_0)$.

Analogously to what done, from (6.3) to (6.8), the equalities (6.9) and (6.10) hold, if we are able to show that

$$E_0^{i-1}[(X_i - X_{i-1})^7] = R(\theta, \Delta_n^4, X_{i-1}), \tag{6.13}$$

$$E_0^{i-1}[(X_i - X_{i-1})^8] = 7 \cdot 5 \cdot 3\Delta_n^4 c_{i-1}^4(\beta_0) + R(\theta, \Delta_n^5, X_{i-1}). \tag{6.14}$$

We only prove (6.14), because (6.13) follows by means of similar arguments. The application of the Itô-Taylor formula to the function $f_x(y) = (y-x)^8$ yields

$$\begin{aligned} E_0^{i-1}\big[(X_i - X_{i-1})^8\big] &= f_{X_{i-1}}(X_{i-1}) + \Delta_n L_0 f_{X_{i-1}}(X_{i-1}) + \frac{\Delta_n^2}{2} L_0^2 f_{X_{i-1}}(X_{i-1}) \\ &\quad + \frac{\Delta_n^3}{3!} L_0^3 f_{X_{i-1}}(X_{i-1}) + \frac{\Delta_n^4}{4!} L_0^4 f_{X_{i-1}}(X_{i-1}) \\ &\quad + \int_0^{\Delta_n}\int_0^{u_1}\int_0^{u_2}\int_0^{u_3}\int_0^{u_4} E_0^{i-1}\big[L_0^5 f_{X_{i-1}}(X_{t_{i-1}^n+u_5})\big]\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_3\mathrm{d}u_4\mathrm{d}u_5 \end{aligned}$$

By applying (6.2), we get

$$\int_0^{\Delta_n}\int_0^{u_1}\int_0^{u_2}\int_0^{u_3}\int_0^{u_4} E_0^{i-1}\big[L_0^5 f_{X_{i-1}}(X_{t_{i-1}^n+u_5})\big]\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_3\mathrm{d}u_4\mathrm{d}u_5 = R(\theta, \Delta_n^5, X_{i-1}).$$

Furthermore, by means of long and cumbersome calculations, we can show that $f_x(x) = L_0 f_x(x) = L_0^2 f_x(x) = L_0^3 f_x(x) = 0$ while $L_0^4 f_x(x) = 7 \cdot 5 \cdot 3 \cdot 4! c_{i-1}^4(\beta_0)$. □

**Lemma 3** (Triangular arrays convegence) *Let $U_i^n$ and $U$ be random variables, with $U_i^n$ being $\mathcal{G}_i^n$-measurable. The two following conditions imply $\sum_{i=1}^n U_i^n \xrightarrow[n\to\infty]{P} U$:*

$$\sum_{i=1}^n E[U_i^n|\mathcal{G}_{i-1}^n] \xrightarrow[n\to\infty]{P} U, \quad \sum_{i=1}^n E[(U_i^n)^2|\mathcal{G}_{i-1}^n] \xrightarrow[n\to\infty]{P} 0$$

*Proof* See the proof of Lemma 9 in Genon-Catalot and Jacod (1993). □

**Lemma 4** *Let $f : \Theta \times \mathbb{R} \to \mathbb{R}$ be such that $f(\theta, x) \in C_{\uparrow}^{1,1}(\Theta \times \mathbb{R}, \mathbb{R})$. Let us assume $A_1 - A_6$, if $\Delta_n \to 0$ and $n\Delta_n \to \infty$ we have that*

$$\frac{1}{n}\sum_{i=1}^n f_{i-1}(\theta) \xrightarrow[n\to\infty]{P_0} \int f(x,\theta)\pi_0(\mathrm{d}x)$$

*uniformly in $\theta$.*

*Proof* See the proof of Lemma 8 in Kessler (1997). □

**Lemma 5** *Let $f : \Theta \times \mathbb{R} \to \mathbb{R}$ be such that $f(\theta, x) \in C_{\uparrow}^{1,1}(\Theta \times \mathbb{R}, \mathbb{R})$. Let us assume $A_1 - A_6$, if $\Delta_n \to 0$ and $n\Delta_n \to \infty$, as $n \to \infty$, we have that*

$$\frac{1}{n\Delta_n^j}\sum_{i=1}^n f_{i-1}(\theta)(X_i - r_l(\Delta_n, X_{i-1}, \theta_0))^k \xrightarrow[n\to\infty]{P_0} \begin{cases} 0, & j=1, k=1, \\ \int f(\theta,x)c(\beta_0,x)\pi_0(\mathrm{d}x), & j=1, k=2, \\ \int f(\theta,x)R(\theta,1,x)\pi_0(\mathrm{d}x), & j=2, k=3, \\ 0, & j=1, k=4, \\ 3\int f(\theta,x)c^2(\beta_0,x)\pi_0(\mathrm{d}x), & j=2, k=4, \end{cases}$$

*uniformly in $\theta$.*

*Proof* The cases $j = 1, k = 1$ and $j = 1, k = 2$ coincide with Lemma 9 and Lemma 10 in Kessler (1997) and then we use the same approach to show that remaining convergences hold true.

By setting

$$\zeta_i^n(\theta) := \frac{1}{n\Delta_n^2} f_{i-1}(\theta)(X_i - r_l(\Delta_n, X_{i-1}, \theta_0))^3,$$

we prove that the convergence holds for all $\theta$. By taking into account Lemma 2 and Lemma 4

$$E_0^{i-1}\left[\zeta_i^n(\theta)\right] = \frac{1}{n}\sum_{i=1}^n f_{i-1}(\theta)R(\theta,1,X_{i-1}) \xrightarrow[n\to\infty]{P_0} \int f(\theta,x)R(\theta,1,x)\pi_0(\mathrm{d}x),$$

$$E_0^{i-1}\left[\left(\zeta_i^n(\theta)\right)^2\right] = \frac{1}{n^2\Delta_n}\sum_{i=1}^n [5\cdot 3c_{i-1}^3(\beta_0) + R(\theta,1,X_{i-1})] \xrightarrow[n\to\infty]{P_0} 0.$$

Therefore by Lemma 3 we can conclude that

$$\zeta_i^n(\theta) \xrightarrow[n\to\infty]{P_0} \int f(\theta,x)R(\theta,1,x)\pi_0(\mathrm{d}x),$$

for all $\theta$. For the uniformity it is sufficient to prove the tightness of the sequence of random elements

$$Y_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} \frac{f_{i-1}(\theta)(X_i - r_l(\Delta_n, X_{i-1}, \theta_0))^3}{\Delta_n^2}$$

taking values in the Banach space $C(\Theta)$ endowed with the sup-norm $||\cdot||_\infty$. From the assumptions of lemma follows that $\sup_n E_0[\sup_{\theta\in\Theta}|\partial_\theta Y_n(\theta)|] < \infty$ which implies the tightness of $Y_n(\theta)$ for the criterion given by Theorem 16.5 in Kallenberg (2001).

By setting

$$\zeta_i^n(\theta) := \frac{1}{n\Delta_n^2} f_{i-1}(\theta)(X_i - r_l(\Delta_n, X_{i-1}, \theta_0))^4,$$

we prove that the convergence holds for all $\theta$. By taking into account Lemmas 2 and 4

$$E_0^{i-1}\left[\zeta_i^n(\theta)\right] = \frac{1}{n}\sum_{i=1}^{n} f_{i-1}(\theta)\left[3c_{i-1}^2(\beta_0) + R(\theta, \Delta_n, X_{i-1})\right] \xrightarrow[n\to\infty]{P_0} 3\int f(\theta, x)c^2(\beta_0, x)\pi_0(\mathrm{d}x),$$

$$E_0^{i-1}\left[\left(\zeta_i^n(\theta)\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n}\left[7\cdot 5\cdot 3c_{i-1}^4(\beta_0) + R(\theta, \Delta_n, X_{i-1})\right] \xrightarrow[n\to\infty]{P_0} 0.$$

Therefore by Lemma 3 we get the pointwise convergence. For the uniformity of the convergence we proceed as done above. □

Before to proceed with the proofs of the main theorems of the paper, we introduce some useful quantities coinciding with (4.2)–(4.8) appearing in Kessler (1997). We can write down

$$\mathtt{l}_{p,i}(\theta) - \mathtt{l}_{p,i}(\theta_0) = \varphi_{i,1}(\theta,\theta_0) + \varphi_{i,2}(\theta,\theta_0) + \varphi_{i,3}(\theta,\theta_0) + \varphi_{i,4}(\theta,\theta_0), \tag{6.15}$$

where

$$\varphi_{i,1}(\theta,\theta_0) := \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))^2}{2\Delta_n}\left\{\frac{1+\sum_{j=1}^{k_0}\Delta_n^j \mathrm{d}_j(\theta, X_{i-1})}{c_{i-1}(\beta)} - \frac{1+\sum_{j=1}^{k_0}\Delta_n^j \mathrm{d}_j(\theta_0, X_{i-1})}{c_{i-1}(\beta_0)}\right\},$$

$$\varphi_{i,2}(\theta,\theta_0) := \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))(r_{k_0}(\Delta_n, X_{i-1}, \theta_0) - r_{k_0}(\Delta_n, X_{i-1}, \theta))}{\Delta_n c_{i-1}(\beta)} \times\left\{1+\sum_{j=1}^{k_0}\Delta_n^j \mathrm{d}_j(\theta, X_{i-1})\right\},$$

$$\varphi_{i,3}(\theta,\theta_0) := \frac{(r_{k_0}(\Delta_n, X_{i-1}, \theta_0) - r_{k_0}(\Delta_n, X_{i-1}, \theta))^2}{2\Delta_n c_{i-1}(\beta)}\left\{1+\sum_{j=1}^{k_0}\Delta_n^j \mathrm{d}_j(\theta, X_{i-1})\right\},$$

$$\varphi_{i,4}(\theta,\theta_0) := \frac{1}{2}\log\left(\frac{c_{i-1}(\beta)}{c_{i-1}(\beta_0)}\right) + \frac{1}{2}\sum_{j=1}^{k_0}\Delta_n^j(\mathrm{e}_j(\theta, X_{i-1}) - \mathrm{e}_j(\theta_0, X_{i-1})).$$

Furthermore

$$\partial_{\alpha_h} \mathbb{l}_{p,i}(\theta) = \eta^h_{i,1}(\theta) + \eta^h_{i,2}(\theta), \quad h = 1, 2, \dots, m_1, \tag{6.16}$$

where

$$\eta^h_{i,1}(\theta) := -(\partial_{\alpha_h} r_{k_0}(\Delta_n, X_{i-1}, \theta))(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta)) \frac{\left\{1 + \sum_{j=1}^{k_0} \Delta_n^j \mathrm{d}_j(\theta, X_{i-1})\right\}}{\Delta_n c_{i-1}(\beta)},$$

$$\eta^h_{i,2}(\theta) := (X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta))^2 \frac{\sum_{j=1}^{k_0} \Delta_n^j \partial_{\alpha_h} \mathrm{d}_j(\theta, X_{i-1})}{2\Delta_n c_{i-1}(\beta)} + \frac{1}{2} \sum_{j=1}^{k_0} \Delta_n^j \partial_{\alpha_h} \mathrm{e}_j(\theta, X_{i-1}),$$

and

$$\partial_{\beta_k} \mathbb{l}_{p,i}(\theta) = \xi^k_{i,1}(\theta) + \xi^k_{i,2}(\theta) + \xi^k_{i,3}(\theta), \quad k = 1, 2, \dots, m_2, \tag{6.17}$$

where

$$\xi^k_{i,1}(\theta) := \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta))^2}{2\Delta_n c_{i-1}(\beta)} \left\{ \sum_{j=1}^{k_0} \Delta_n^j \partial_{\beta_k} \mathrm{d}_j(\theta, X_{i-1}) \right\} + \frac{1}{2} \sum_{j=1}^{k_0} \Delta_n^j \partial_{\beta_k} \mathrm{e}_j(\theta, X_{i-1}),$$

$$\xi^k_{i,2}(\theta) := -\frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta))^2 \partial_{\beta_k} c_{i-1}(\beta)}{2\Delta_n c^2_{i-1}(\beta)} \left\{ 1 + \sum_{j=1}^{k_0} \Delta_n^j \mathrm{d}_j(\theta, X_{i-1}) \right\} + \frac{\partial_{\beta_k} c_{i-1}(\beta)}{2 c_{i-1}(\beta)},$$

$$\xi^k_{i,3}(\theta) := -(\partial_{\beta_k} r_{k_0}(\Delta_n, X_{i-1}, \theta))(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta)) \frac{\left\{1 + \sum_{j=1}^{k_0} \Delta_n^j \mathrm{d}_j(\theta, X_{i-1})\right\}}{\Delta_n c_{i-1}(\beta)}.$$

From (6.15) it is possible to derive

$$\partial^2_{\alpha_h \alpha_k} \mathbb{l}_{p,i}(\theta) := \delta^{h,k}_{i,1}(\theta) + \delta^{h,k}_{i,2}(\theta) + \delta^{h,k}_{i,3}(\theta) + \delta^{h,k}_{i,4}(\theta), \quad h, k = 1, 2, \dots, m_1, \tag{6.18}$$

where

$$\delta^{h,k}_{i,1}(\theta) := \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))^2}{2 c_{i-1}(\beta)} \left\{ \left( \partial^2_{\alpha_h \alpha_k} \mathrm{d}_1 \right)_{i-1}(\theta) + R(\theta, \Delta_n, X_{i-1}) \right\},$$

$$\delta^{h,k}_{i,2}(\theta) := \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))}{c_{i-1}(\beta)} \left\{ -\partial^2_{\alpha_h \alpha_k} b_{i-1}(\alpha) + R(\theta, \Delta_n, X_{i-1}) \right\},$$

$$\delta^{h,k}_{i,3}(\theta) := \frac{1}{2} \Delta_n \partial^2_{\alpha_h \alpha_k} \mathrm{e}_1(\theta, X_{i-1}),$$

$$\delta^{h,k}_{i,4}(\theta) := \Delta_n \left\{ \frac{\partial^2_{\alpha_h \alpha_k} b_{i-1}(\alpha)(b_{i-1}(\alpha) - b_{i-1}(\alpha_0)) + \partial_{\alpha_h} b_{i-1}(\alpha) \partial_{\alpha_k} b_{i-1}(\alpha)}{c_{i-1}(\beta)} \right.$$
$$\left. + R(\theta, \Delta_n, X_{i-1}) \right\},$$

$$\partial^2_{\beta_h \beta_k} \mathbb{l}_{p,i}(\theta) := \nu^{h,k}_{i,1}(\theta) + \nu^{h,k}_{i,2}(\theta) + \nu^{h,k}_{i,3}(\theta), \quad h, k = 1, 2, \dots, m_2, \tag{6.19}$$

where

$$\nu^{h,k}_{i,1}(\theta) := \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))^2}{2\Delta_n} \left\{ \left( \partial^2_{\beta_h \beta_k} c^{-1} \right)_{i-1}(\beta) + R(\theta, \Delta_n, X_{i-1}) \right\},$$

$$\nu^{h,k}_{i,2}(\theta) := \frac{1}{2}(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0)) R(\theta, 1, X_{i-1})),$$

$$\nu^{h,k}_{i,3}(\theta) := \frac{1}{2}(\partial^2_{\beta_h \beta_k} \log c)_{i-1}(\beta) + R(\theta, \Delta_n, X_{i-1})),$$

and

$$\partial^2_{\alpha_h \beta_k} \daleth_{p,i}(\theta) := \mu_{i,1}(\theta) + \mu_{i,2}(\theta), \quad h = 1, 2, \ldots, m_1, k = 1, 2, \ldots, m_2, \tag{6.20}$$

where

$$\mu_{i,1}(\theta) := \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))^2}{2\Delta_n} R(\theta, \Delta_n, X_{i-1}),$$
$$\mu_{i,2}(\theta) := \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))}{\Delta_n} R(\theta, \Delta_n, X_{i-1}) + R(\theta, \Delta_n, X_{i-1}).$$

*Proof of Theorem 2* We observe that

$$\mathbb{D}_{p,n}(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{4} (\varphi_{i,k}(\theta, \theta_0))^2 + 2 \sum_{j<k} )\varphi_{i,j}(\theta, \theta_0) \varphi_{i,k}(\theta, \theta_0) \right\}.$$

Under $H_0$, from Lemmas 2 and 5, we derive

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} (\varphi_{i,1}(\theta, \theta_0))^2 &= \frac{1}{n} \sum_{i=1}^{n} \Bigg[ \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))^4}{4\Delta_n^2} \\
&\qquad \left\{ \frac{1}{c_{i-1}(\beta)} - \frac{1}{c_{i-1}(\beta_0)} + R(\theta, \Delta_n, X_{i-1}) \right\}^2 \Bigg] \\
&= \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))^4}{4\Delta_n^2} \left\{ \frac{1}{c_{i-1}(\beta)} - \frac{1}{c_{i-1}(\beta_0)} \right\}^2 \right] \\
&\quad + \mathbf{o}_{P_0}(1) \xrightarrow[n \to \infty]{P_0} \frac{3}{4} \int c^2(\beta_0, x) \left\{ \frac{1}{c(\beta, x)} - \frac{1}{c(\beta_0, x)} \right\}^2 \pi_0(\mathrm{d}x) \\
\frac{1}{n} \sum_{i=1}^{n} (\varphi_{i,2}(\theta, \theta_0))^2 &= \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))^2}{c_{i-1}^2(\beta_0)} [b_{i-1}(\alpha_0) - b_{i-1}(\alpha)]^2 \right] + \mathbf{o}_{P_0}(1) \\
&\xrightarrow[n \to \infty]{P_0} 0 \\
\frac{1}{n} \sum_{i=1}^{n} (\varphi_{i,3}(\theta, \theta_0))^2 &= \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\Delta_n^2 [b_{i-1}(\alpha_0) - b_{i-1}(\alpha)]^4}{4c_{i-1}^2(\beta)} \right] + \mathbf{o}_{P_0}(1) \\
&\xrightarrow[n \to \infty]{P_0} 0 \\
\frac{1}{n} \sum_{i=1}^{n} (\varphi_{i,4}(\theta, \theta_0))^2 &= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{4} \left[ \log \left( \frac{c_{i-1}(\beta)}{c_{i-1}(\beta_0)} \right) \right]^2 + \mathbf{o}_{P_0}(1) \\
&\xrightarrow[n \to \infty]{P_0} \frac{1}{4} \int \left[ \log \left( \frac{c(\beta, x)}{c(\beta_0, x)} \right) \right]^2 \pi_0(\mathrm{d}x) \\
\frac{1}{n} \sum_{i=1}^{n} \varphi_{i,1}(\theta, \theta_0) \varphi_{i,4}(\theta, \theta_0) &= \frac{1}{n} \sum_{i=1}^{n} \frac{(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta_0))^2}{4\Delta_n} \left\{ \frac{1}{c_{i-1}(\beta)} - \frac{1}{c_{i-1}(\beta_0)} \right\} \\
&\qquad \times \log \left( \frac{c_{i-1}(\beta)}{c_{i-1}(\beta_0)} \right) + \mathbf{o}_{P_0}(1)
\end{aligned}$$

$$\xrightarrow[n\to\infty]{P_0} \frac{1}{4}\int c(\beta_0,x)\left\{\frac{1}{c(\beta,x)}-\frac{1}{c(\beta_0,x)}\right\}\log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)\pi_0(\mathrm{d}x)$$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_{i,1}(\theta,\theta_0)\varphi_{i,j}(\theta,\theta_0) \xrightarrow[n\to\infty]{P_0} 0, \quad j=2,3,$$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_{i,2}(\theta,\theta_0)\varphi_{i,j}(\theta,\theta_0) \xrightarrow[n\to\infty]{P_0} 0, \quad j=3,4,$$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_{i,3}(\theta,\theta_0)\varphi_{i,4}(\theta,\theta_0) \xrightarrow[n\to\infty]{P_0} 0,$$

uniformly in $\theta$. Thus the statement of the theorem immediately follows. □

Let

$$C_{p,n}(\theta,\theta_0) := \begin{pmatrix} \frac{1}{n\Delta_n}[\partial^2_{\alpha_h\alpha_k}T_{p,n}(\theta,\theta_0)]_{\substack{h=1,\dots,m_1\\k=1,\dots,m_1}} & \frac{1}{n\sqrt{\Delta_n}}[\partial^2_{\alpha_h\beta_k}T_{p,n}(\theta,\theta_0)]_{\substack{h=1,\dots,m_1\\k=1,\dots,m_2}} \\ \frac{1}{n\sqrt{\Delta_n}}[\partial^2_{\alpha_h\beta_k}T_{p,n}(\theta,\theta_0)]_{\substack{h=1,\dots,m_1\\k=1,\dots,m_2}} & \frac{1}{n}[\partial^2_{\beta_h\beta_k}T_{p,n}(\theta,\theta_0)]_{\substack{h=1,\dots,m_2\\k=1,\dots,m_2}} \end{pmatrix} \tag{6.21}$$

where

$$\partial^2_{\alpha_h\alpha_k}T_{p,n}(\theta,\theta_0) = 2\sum_{i=1}^{n}\left\{\partial_{\alpha_h}\mathbb{1}_{p,i}(\theta)\partial_{\alpha_k}\mathbb{1}_{p,i}(\theta)+[\mathbb{1}_{p,i}(\theta)-\mathbb{1}_{p,i}(\theta_0)]\partial^2_{\alpha_h\alpha_k}\mathbb{1}_{p,i}(\theta)\right\}, \tag{6.22}$$

$$\partial^2_{\beta_h\beta_k}T_{p,n}(\theta,\theta_0) = 2\sum_{i=1}^{n}\left\{\partial_{\beta_h}\mathbb{1}_{p,i}(\theta)\partial_{\beta_k}\mathbb{1}_{p,i}(\theta)+[\mathbb{1}_{p,i}(\theta)-\mathbb{1}_{p,i}(\theta_0)]\partial^2_{\beta_h\beta_k}\mathbb{1}_{p,i}(\theta)\right\}, \tag{6.23}$$

$$\partial^2_{\alpha_h\beta_k}T_{p,n}(\theta,\theta_0) = 2\sum_{i=1}^{n}\left\{\partial_{\alpha_h}\mathbb{1}_{p,i}(\theta)\partial_{\beta_k}\mathbb{1}_{p,i}(\theta)+[\mathbb{1}_{p,i}(\theta)-\mathbb{1}_{p,i}(\theta_0)]\partial^2_{\alpha_h\beta_k}\mathbb{1}_{p,i}(\theta)\right\}. \tag{6.24}$$

The following proposition concerning the asymptotic behavior of $C_{p,n}(\theta,\theta_0)$ plays a crucial role in the proof of Theorem 3.

**Proposition 1** *Under $H_0$, assume $A_1 - A_6$ and $\Delta_n \to 0, n\Delta_n \to \infty$, as $n \to \infty$. The following convergences hold*

$$C_{p,n}(\theta_0,\theta_0) \xrightarrow[n\to\infty]{P_0} 2I(\theta_0) \tag{6.25}$$

*and*

$$\sup_{||\theta||\le\varepsilon_n} ||C_{p,n}(\theta_0+\theta,\theta_0)-C_{p,n}(\theta_0,\theta_0)|| \xrightarrow[n\to\infty]{P_0} 0, \quad \varepsilon_n \to 0. \tag{6.26}$$

*Proof of Proposition 1* We study the uniform convergence in probability of $C_{p,n}(\theta,\theta_0)$. Thus we prove that uniformly in $\theta$

$$C_{p,n}(\theta,\theta_0) \xrightarrow[n\to\infty]{P_0} 2K(\theta,\theta_0) := 2\begin{pmatrix} K_1(\theta,\theta_0)+K_2(\theta,\theta_0) & 0 \\ 0 & K_3(\theta,\theta_0)+K_4(\theta,\theta_0)\end{pmatrix} \tag{6.27}$$

where

$$\begin{aligned}
K_1(\theta,\theta_0) &:= \int \frac{\partial_{\alpha_h} b(\alpha,x)\partial_{\alpha_k} b(\alpha,x)}{c^2(\beta,x)} c(\beta_0,x)\pi_0(\mathrm{d}x),\\
K_2(\theta,\theta_0) &:= \frac{1}{4}\int \partial^2_{\alpha_h\alpha_k} \mathrm{d}_1(x,\theta)\left[\frac{c(\beta_0,x)}{c(\beta,x)}-1\right]\left[3\frac{c(\beta_0,x)}{c(\beta,x)}+\log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)-1\right]\pi_0(\mathrm{d}x)\\
&\quad+\frac{1}{2}\int\left[\frac{\partial^2_{\alpha_h\alpha_k} b(\alpha,x)(b(\alpha,x)-b(\alpha_0,x))+\partial_{\alpha_h} b(\alpha,x)\partial_{\alpha_k} b(\alpha,x)}{c(\beta,x)}\right]\\
&\quad\times\left[\frac{c(\beta_0,x)}{c(\beta,x)}-1+\log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)\right]\pi_0(\mathrm{d}x)\\
&\quad+\int\frac{-\partial^2_{\alpha_h\alpha_k} b(\alpha,x)}{c(\beta,x)}\\
&\quad\times\left[\frac{1}{2}\left(\frac{1}{c(\beta_0,x)}-\frac{1}{c(\beta,x)}\right)R(\theta,1,x)+\frac{c(\beta_0,x)}{c^2(\beta,x)}(b(\alpha,x)-b(\alpha_0,x))\right]\pi_0(\mathrm{d}x)\\
K_3(\theta,\theta_0) &:= \frac{1}{2}\int\left\{\frac{c(\beta_0,x)\partial_{\beta_h} c(\beta,x)\partial_{\beta_k} c(\beta,x)}{c^3(\beta,x)}\left[\frac{3}{2}\frac{c(\beta_0,x)}{c(\beta,x)}-1\right]\right.\\
&\quad\left.+\frac{1}{2}\frac{\partial_{\beta_h} c(\beta,x)\partial_{\beta_k} c(\beta,x)}{c^2(\beta,x)}\right\}\pi_0(\mathrm{d}x),\\
K_4(\theta,\theta_0) &:= \frac{1}{4}\int c(\beta_0,x)\partial^2_{\beta_h\beta_k}\log c(\beta,x)\left[\frac{1}{c(\beta,x)}-\frac{1}{c(\beta_0,x)}\right]\pi_0(\mathrm{d}x)\\
&\quad+\frac{1}{4}\int\log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)c(\beta_0,x)\partial^2_{\beta_h\beta_k} c^{-1}(\beta,x)\pi_0(\mathrm{d}x)\\
&\quad+\frac{1}{4}\int\log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)\partial^2_{\beta_h\beta_k}\log c(\beta,x)\pi_0(\mathrm{d}x).
\end{aligned}$$

Let us start with the analysis of the quantity $\frac{1}{n\Delta_n}\partial^2_{\alpha_h\alpha_k} T_{p,n}(\theta,\theta_0)$ given by (6.22) which can be split in two terms. From (6.16) folllows that

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \partial_{\alpha_h} \mathrm{l}_{p,i}(\theta)\partial_{\alpha_k} \mathrm{l}_{p,i}(\theta) = \frac{1}{n\Delta_n}\sum_{i=1}^n (\eta^h_{i,1}(\theta)+\eta^h_{i,2}(\theta))(\eta^k_{i,1}(\theta)+\eta^k_{i,2}(\theta))$$

for each $\theta\in\Theta$. Since $\partial_{\alpha_h} r_{k_0}(\Delta_n, X_{i-1},\theta) = \Delta_n\partial_{\alpha_h} b_{i-1}(\alpha) + R(\theta,\Delta_n^2, X_{i-1})$, by taking into account Lemma 5, we get

$$\begin{aligned}
\frac{1}{n\Delta_n}\sum_{i=1}^n \partial_{\alpha_h} \mathrm{l}_{p,i}(\theta)\partial_{\alpha_k} \mathrm{l}_{p,i}(\theta) &= \frac{1}{n\Delta_n}\sum_{i=1}^n \eta^h_{i,1}(\theta)\eta^k_{i,1}(\theta) + \mathbf{o}_{P_0}(1)\\
&= \frac{1}{n\Delta_n}\sum_{i=1}^n \frac{\partial_{\alpha_h} b_{i-1}(\alpha)\partial_{\alpha_k} b_{i-1}(\alpha)}{c^2_{i-1}(\beta)}
\end{aligned}$$

$$(X_i - r_{k_0}(\Delta_n, X_{i-1}, \theta))^2 + \mathbf{o}_{P_0}(1) \xrightarrow[n\to\infty]{P_0} K_1(\theta, \theta_0) \tag{6.28}$$

uniformly in $\theta$. Now, by resorting (6.15) and (6.18), we rewrite the second term appearing in (6.22) as follows

$$\frac{1}{n\Delta_n}\sum_{i=1}^n [\mathbb{1}_{p,i}(\theta) - \mathbb{1}_{p,i}(\theta_0)]\partial^2_{\alpha_h\alpha_k}\mathbb{1}_{p,i}(\theta) = \frac{1}{n\Delta_n}\sum_{i=1}^n \left[\sum_{l=1}^4\sum_{j=1}^4 \varphi_{i,l}(\theta,\theta_0)\delta_{i,j}^{h,k}(\theta)\right].$$

By applying Lemmas 1 and 5, the following convergence results hold

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,1}(\theta,\theta_0)\delta_{i,1}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} \frac{3}{4}\int \partial^2_{\alpha_h\alpha_k}\mathrm{d}_1(\theta,x)\frac{c^2(\beta_0,x)}{c(\beta,x)}\left[\frac{1}{c(\beta,x)} - \frac{1}{c(\beta_0,x)}\right]\pi_0(\mathrm{d}x),$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,1}(\theta,\theta_0)\delta_{i,2}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} \frac{1}{2}\int \frac{-\partial^2_{\alpha_h\alpha_k}b(\alpha,x)}{c(\beta,x)}\left[\frac{1}{c(\beta,x)} - \frac{1}{c(\beta_0,x)}\right]R(\theta,1,x)\pi_0(\mathrm{d}x),$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,1}(\theta,\theta_0)\delta_{i,3}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} \frac{1}{4}\int \partial^2_{\alpha_h\alpha_k}\mathrm{e}_1(\theta,x)\left[\frac{c(\beta_0,x)}{c(\beta,x)} - 1\right]\pi_0(\mathrm{d}x),$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,1}(\theta,\theta_0)\delta_{i,4}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} \frac{1}{2}\int\left[\frac{c(\beta_0,x)}{c(\beta,x)} - 1\right]\left[\frac{\partial^2_{\alpha_h\alpha_k}b(\alpha,x)(b(\alpha,x)-b(\alpha_0,x)) + \partial_{\alpha_h}b(\alpha,x)\partial_{\alpha_k}b(\alpha,x)}{c(\beta,x)}\right]\pi_0(\mathrm{d}x),$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,2}(\theta,\theta_0)\delta_{i,2}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} \int \frac{c(\beta_0,x)}{c^2(\beta,x)}(-\partial^2_{\alpha_h\alpha_k}b(\alpha,x))(b(\alpha,x)-b(\alpha_0,x))\pi_0(\mathrm{d}x),$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,4}(\theta,\theta_0)\delta_{i,1}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} \frac{1}{4}\int \log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)\frac{c(\beta_0,x)}{c(\beta,x)}\partial^2_{\alpha_h\alpha_k}\mathrm{d}_1(\theta,x)\pi_0(\mathrm{d}x),$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,4}(\theta,\theta_0)\delta_{i,3}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} \frac{1}{4}\int \partial^2_{\alpha_h\alpha_k}\mathrm{e}_1(\theta,x)\log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)\pi_0(\mathrm{d}x),$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,4}(\theta,\theta_0)\delta_{i,4}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} \frac{1}{2}\int \log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)\left\{\frac{\partial^2_{\alpha_h\alpha_k}b(\alpha,x)(b(\alpha,x)-b(\alpha_0,x)) + \partial_{\alpha_k}b(\alpha,x)\partial_{\alpha_h}b(\alpha,x)}{c(\beta,x)}\right\}\pi_0(\mathrm{d}x),$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,2}(\theta,\theta_0)\delta_{i,j}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} 0, \quad j = 1,3,4,$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,3}(\theta,\theta_0)\delta_{i,j}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} 0, \quad j=1,2,3,4,$$

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \varphi_{i,4}(\theta,\theta_0)\delta_{i,2}^{h,k}(\theta) \xrightarrow[n\to\infty]{P_0} 0,$$

uniformly in $\theta$. Finally, since $\mathrm{d}_1(\theta,x)=-\mathrm{e}_1(\theta,x)$, we get

$$\frac{1}{n\Delta_n}\sum_{i=1}^n [\mathbb{l}_{p,i}(\theta)-\mathbb{l}_{p,i}(\theta_0)]\partial^2_{\alpha_h\alpha_k}\mathbb{l}_{p,i}(\theta) \xrightarrow[n\to\infty]{P_0} K_2(\theta,\theta_0). \tag{6.29}$$

uniformly in $\theta$. Hence, by (6.28) and (6.29), we immediately derive

$$\frac{1}{n\Delta_n}\partial^2_{\alpha_h\alpha_k}T_{p,n}(\theta,\theta_0) \xrightarrow[n\to\infty]{P_0} 2(K_1(\theta,\theta_0)+K_2(\theta,\theta_0)) \tag{6.30}$$

uniformly in $\theta$.

Now, we consider the elements of the matrix $C_{n,p}(\theta,\theta_0)$ given by (6.23). First, we study the convergence probability of

$$\frac{1}{n}\sum_{i=1}^n \partial_{\beta_h}\mathbb{l}_{p,i}(\theta)\partial_{\beta_k}\mathbb{l}_{p,i}(\theta)=\frac{1}{n}\sum_{i=1}^n(\xi_{i,1}^h(\theta)+\xi_{i,2}^h(\theta)+\xi_{i,3}^h(\theta))(\xi_{i,1}^k(\theta)+\xi_{i,2}^k(\theta)+\xi_{i,3}^k(\theta)).$$

Since $\partial_{\beta_h}r_{k_0}(\Delta_n,X_{i-1},\theta)=R(\theta,\Delta_n^2,X_{i-1})$, from Lemmas 5 and 1 we derive

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \partial_{\beta_h}\mathbb{l}_{p,i}(\theta)\partial_{\beta_k}\mathbb{l}_{p,i}(\theta) &= \frac{1}{n}\sum_{i=1}^n \xi_{i,2}^h(\theta)\xi_{i,2}^k(\theta)+\mathbf{o}_{P_0}(1)\\
&= \frac{1}{n}\sum_{i=1}^n \frac{\partial_{\beta_h}c_{i-1}(\beta)\partial_{\beta_k}c_{i-1}(\beta)}{4\Delta_n^2 c_{i-1}^4(\beta)}(X_i-r_{k_0}(\Delta_n,X_{i-1},\theta))^4\\
&\quad+\frac{1}{n}\sum_{i=1}^n \frac{\partial_{\beta_h}c_{i-1}(\beta)\partial_{\beta_k}c_{i-1}(\beta)}{2\Delta_n c_{i-1}^3(\beta)}(X_i-r_{k_0}(\Delta_n,X_{i-1},\theta))^2\\
&\quad+\frac{1}{n}\sum_{i=1}^n \frac{\partial_{\beta_h}c_{i-1}(\beta)\partial_{\beta_k}c_{i-1}(\beta)}{4c_{i-1}^2(\beta)}+\mathbf{o}_{P_0}(1)\\
&\xrightarrow[n\to\infty]{P_0} K_3(\theta,\theta_0)
\end{aligned} \tag{6.31}$$

uniformly in $\theta$. Now, by resorting (6.15) and (6.19), we rewrite the second term appearing in (6.23) as follows

$$\frac{1}{n}\sum_{i=1}^n [\mathbb{l}_{p,i}(\theta)-\mathbb{l}_{p,i}(\theta_0)]\partial^2_{\beta_h\beta_k}\mathbb{l}_{p,i}(\theta)=\frac{1}{n}\sum_{i=1}^n\left[\sum_{k=1}^4\sum_{j=1}^3\varphi_{i,k}(\theta,\theta_0)\nu_{i,j}^{h,k}(\theta)\right].$$

By taking into account again Lemmas 1 and 5, the following results yield

$$\frac{1}{n}\sum_{i=1}^n \varphi_{i,1}(\theta,\theta_0)\nu_{i,3}(\theta) \xrightarrow[n\to\infty]{P_0} \frac{1}{4}\int c(\beta_0,x)\partial^2_{\beta_h\beta_k}\log c(\beta,x)\left[\frac{1}{c(\beta,x)}-\frac{1}{c(\beta_0,x)}\right]\pi_0(\mathrm{d}x)$$

$$\frac{1}{n}\sum_{i=1}^n \varphi_{i,4}(\theta,\theta_0)\nu_{i,1}(\theta) \xrightarrow[n\to\infty]{P_0} \frac{1}{4}\int \log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)c(\beta_0,x)\partial^2_{\beta_h\beta_k}c^{-1}(\beta,x)\pi_0(\mathrm{d}x)$$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_{i,4}(\theta,\theta_0)\nu_{i,3}(\theta)\xrightarrow[n\to\infty]{P_0}\frac{1}{4}\int\log\left(\frac{c(\beta,x)}{c(\beta_0,x)}\right)\partial^2_{\beta_h\beta_k}\log c(\beta,x)\pi_0(\mathrm{d}x)$$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_{i,1}(\theta,\theta_0)\nu_{i,j}(\theta)\xrightarrow[n\to\infty]{P_0}0,\quad j=1,2,$$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_{i,k}(\theta,\theta_0)\nu_{i,j}(\theta)\xrightarrow[n\to\infty]{P_0}0,\quad k=2,3,\ j=1,2,3,$$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_{i,4}(\theta,\theta_0)\nu_{i,2}(\theta)\xrightarrow[n\to\infty]{P_0}0,$$

uniformly in $\theta$. Finally

$$\frac{1}{n}\sum_{i=1}^{n}[\mathbb{1}_{p,i}(\theta)-\mathbb{1}_{p,i}(\theta_0)]\partial^2_{\beta_h\beta_k}\mathbb{1}_{p,i}(\theta)\xrightarrow[n\to\infty]{P_0}K_4(\theta,\theta_0)\tag{6.32}$$

uniformly in $\theta$. Therefore, by (6.31) and (6.32), we get

$$\frac{1}{n}\partial^2_{\beta_h\beta_k}T_{p,n}(\theta,\theta_0)\xrightarrow[n\to\infty]{P_0}2(K_3(\theta,\theta_0)+K_4(\theta,\theta_0))\tag{6.33}$$

uniformly in $\theta$.

Recalling the expressions (6.16), (6.17), (6.20) and (6.15), by means of similar arguments adopted above, it is not hard to prove that

$$\frac{1}{n\sqrt{\Delta_n}}\sum_{i=1}^{n}\partial_{\alpha_h}\mathbb{1}_{p,i}(\theta)\partial_{\beta_k}\mathbb{1}_{p,i}(\theta)\xrightarrow[n\to\infty]{P_0}0$$

and

$$\frac{1}{n\sqrt{\Delta_n}}\sum_{i=1}^{n}[\mathbb{1}_{p,i}(\theta)-\mathbb{1}_{p,i}(\theta_0)]\partial^2_{\alpha_h\beta_k}\mathbb{1}_{p,i}(\theta)\xrightarrow[n\to\infty]{P_0}0$$

uniformly in $\theta$. This implies that

$$\frac{1}{n\sqrt{\Delta_n}}\partial^2_{\alpha_h\beta_k}T_{p,n}(\theta,\theta_0)\xrightarrow[n\to\infty]{P_0}0\tag{6.34}$$

uniformly in $\theta$.

In conclusion the results (6.30), (6.33) and (6.34) lead to the convergence (6.27). Moreover, (6.27) implies (6.25) since $K(\theta_0,\theta_0)=I(\theta_0)$. From the inequality

$$\begin{aligned}&\sup_{||\theta||\leq\varepsilon_n}||C_{p,n}(\theta_0+\theta,\theta_0)-C_{p,n}(\theta_0,\theta_0)||\\&\leq\sup_{||\theta||\leq\varepsilon_n}||C_{p,n}(\theta_0+\theta,\theta_0)-2K(\theta_0+\theta,\theta_0)||+\sup_{||\theta||\leq\varepsilon_n}||2K(\theta_0+\theta,\theta_0)-2I(\theta_0)||\\&\quad+||2I(\theta_0)-C_{p,n}(\theta_0,\theta_0)||\end{aligned}$$

follows (6.26). Indeed, (6.25) leads to $||2I(\theta_0)-C_{p,n}(\theta_0,\theta_0)||\xrightarrow[n\to\infty]{P_0}0$, $\varepsilon_n\to 0$, while the term $\sup_{||\theta||\leq\varepsilon_n}||C_{p,n}(\theta_0+\theta,\theta_0)-2K(\theta_0+\theta,\theta_0)||\xrightarrow[n\to\infty]{P_0}0$, $\varepsilon_n\to 0$, by the uniformity of the convergence (i.e. by the result (6.27)). Furthermore, $\sup_{||\theta||\leq\varepsilon_n}||K(\theta_0+\theta,\theta_0)-I(\theta_0)||\xrightarrow[n\to\infty]{P_0}$

0, $\varepsilon_n \to 0$, because the assumptions $A_3$ and $A_5$, imply that $K(\theta, \theta_0)$ is a continuous function with respect to $\theta$. □

Now, we are able to prove Theorem 3.

*Proof of Theorem 3* We adopt classical arguments. By Taylor's formula, we have that

$$\begin{aligned} T_{p,n}(\hat{\theta}_{p,n}, \theta_0) &= T_{p,n}(\theta_0, \theta_0) + n\partial_\theta T_{p,n}(\theta_0, \theta_0)(\hat{\theta}_{p,n} - \theta_0) \\ &\quad + \frac{1}{2}(\varphi(n)^{-1/2}(\hat{\theta}_n - \theta_0))' \Lambda_{p,n}(\hat{\theta}_{p,n}, \theta_0))\varphi(n)^{-1/2}(\hat{\theta}_{p,n} - \theta_0) \\ &= \frac{1}{2}(\varphi(n)^{-1/2}(\hat{\theta}_n - \theta_0))' \Lambda_{p,n}(\hat{\theta}_{p,n}, \theta_0)\varphi(n)^{-1/2}(\hat{\theta}_n - \theta_0) \end{aligned} \tag{6.35}$$

where in the last step we denoted by

$$\begin{aligned} \Lambda_{p,n}(\hat{\theta}_{p,n}, \theta_0) &:= \varphi(n)^{1/2} \int_0^1 (1-u)\partial_\theta^2 T_{p,n}(\theta_0 + u(\hat{\theta}_{p,n} - \theta_0), \theta_0) \mathrm{d}u \varphi(n)^{1/2} \\ &= \int_0^1 (1-u)[C_{p,n}(\theta_0 + u(\hat{\theta}_{p,n} - \theta_0), \theta_0) - C_{p,n}(\theta_0, \theta_0)]\mathrm{d}u + C_{p,n}(\theta_0, \theta_0). \end{aligned}$$

Proposition 1 implies

$$\Lambda_{p,n}(\hat{\theta}_{p,n}, \theta_0) \xrightarrow[n\to\infty]{P_0} 2I(\theta_0). \tag{6.36}$$

By taking into account (6.35), (3.16) and (6.36), Slutsky's theorem allows to conclude the proof. □

*Proof of Theorem 4* Under $H_{1,n}$ we have that [see Lemma 2 in Kitagawa and Uchida (2014)]

$$\varphi(n)^{-1/2}(\hat{\theta}_{p,n} - (\theta_0 + \varphi(n)^{1/2}h)) \xrightarrow[n\to\infty]{d} N(0, I(\theta_0)^{-1}).$$

Therefore, under the hypothesis $H_{1,n}$

$$\varphi(n)^{-1/2}(\hat{\theta}_{p,n} - \theta_0) = \varphi(n)^{-1/2}(\hat{\theta}_{p,n} - \theta) + h \xrightarrow[n\to\infty]{d} N(h, I(\theta_0)^{-1})$$

and

$$C_{p,n}(\hat{\theta}_{p,n}, \theta_0) \xrightarrow[n\to\infty]{P_\theta} 2I(\theta_0) \quad (\text{under } H_{1,n}).$$

Hence, from (6.35) we obtain the result (4.7). □

**Acknowledgements** We would like to thank both the referees for their comments which have greatly improved the first version of the manuscript.

**Compliance with ethical standards**

**Conflicts of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

Aït-Sahalia Y (1996) Testing continuous-time models of the spot interest rate. Rev Financial Stud 70:385–426

Aït-Sahalia Y (2002) Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. Econometrica 70:223–262

Aït-Sahalia Y (2008) Closed-form likelihood expansions for multivariate diffusions. Ann Stat 36:906–937
Balakrishnan V, Sanghvi LD (1968) Distance between populations on the basis of attribute data. Biometrica 24:859–865
Bibby BM, Sørensen M (1995) Martingale estimating functions for discretely observed diffusion processes. Bernoulli 1:17–39
De Gregorio A, Iacus SM (2008) Least squares volatility change point estimation for partially observed diffusion processes. Commun Stat Theory Methods 37:2342–2357
De Gregorio A, Iacus SM (2010) Clustering of discretely observed diffusion processes. Comput Stat Data Anal 54:598–606
De Gregorio A, Iacus SM (2012) Adaptive LASSO-type estimation for multivariate diffusion processes. Econ Theory 28:838–860
De Gregorio A, Iacus SM (2013) On a family of test statistics for discretely observed diffusion processes. J Multivar Anal 122:292–316
Florens-Zmirou D (1989) Approximate discrete-time schemes for statistics of diffusion processes. Statistics 20:547–557
Genon-Catalot V, Jacod J (1993) On the estimation of the diffusion coefficient for multidimensional diffusion processes. Ann Inst Henri Poincaré 29:119–151
Gobet E (2002) LAN property for ergodic diffusions with discrete observations. Ann I H Poincaré-PR 38:711–737
Iacus SM (2008) Simulation and inference for stochastic differential equations: with R examples. Springer series in statistics. Springer, New York
Iacus SM (2011) Option pricing and estimation of financial models with R. Wiley, New York
Iacus SM, Yoshida N (2012) Estimation for the change point of volatility in a stochastic differential equation. Stoch Process Appl 122:1068–1092
Iacus SM, Yoshida N (2017) Simulation and inference for stochastic processes with YUIMA. Springer series in statistics. Springer, New York
Iacus SM, Uchida M, Yoshida N (2009) Parametric estimation for partially hidden diffusion processes sampled at discrete times. Stoch Process Appl 119:1580–1600
Jacod J (2006) Parametric inference for discretely observed non-ergodic diffusions. Bernoulli 12:383–401
Kallenberg O (2001) Foundations of modern probability. Springer, London
Kamatani K, Uchida M (2015) Hybrid multi-step estimators for stochastic differential equations based on sampled data. Stat Inference Stoch Process 18:177–204
Kitagawa H, Uchida M (2014) Adaptive test statistics for ergodic diffusion processes sampled at discrete times. J Stat Plan Inference 150:84–110
Kessler M (1995) Estimation des parametres d'une diffusion par des contrastes corrigés. C R Acad Sci Paris Ser I Math 320:359–362
Kessler M (1997) Estimation of an ergodic diffusion from discrete observations. Scand J Stat 24:211–229
Kessler M, Sørensen M (1999) Estimating equations based on eigenfunctions for a discretely observed diffusion process. Bernoulli 5:299–314
Kutoyants YA (2004) Statistical inference for ergodic diffusion processes. Springer, London
Li C (2013) Maximum-likelihood estimation for diffusion processes via closed-form density expansions. Ann Stat 41:1350–1380
Morales D, Pardo L, Vajda I (1997) Some new statistics for testing hypotheses in parametric models. J Multivar Anal 67:137–168
Pardo L (2006) Statistical inference based on divergence measures. Chapman & Hall/CRC, London
Phillips PCB, Yu J (2009) A two-stage realized volatility approach to estimation of diffusion processes with discrete data. J Econ 150:139–150
Uchida M, Yoshida N (2012) Adaptive estimation of an ergodic diffusion process based on sampled data. Stoch Process Appl 122:2885–2924
Uchida M, Yoshida N (2014) Adaptive Bayes type estimators of ergodic diffusion processes from discrete observations. Stat Inference Stoch Process 17:181–219
Van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge
Yoshida N (1992) Estimation for diffusion processes from discrete observation. J Multivar Anal 41:220–242
Yoshida N (2011) Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic dierential equations. Ann Inst Stat Math 63:431–479