Volume 140   Issue 7   July 2010   ISSN 0378-3758

ELSEVIER

journal of statistical planning and inference

For Contents see page iii

Available online at www.sciencedirect.com

ScienceDirect

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Divergences test statistics for discretely observed diffusion processes

Alessandro De Gregorio [a], Stefano M. Iacus [b,*]

[a] *Department of Statistics, Probability and Applied Statistics, P.le Aldo Moro 5, 00185 Rome, Italy*
[b] *Department of Economics, Business and Statistics, Via Conservatorio 7, 20124 Milan, Italy*

ARTICLE INFO

ABSTRACT

In this paper we propose the use of ϕ-divergences as test statistics to verify simple hypotheses about a one-dimensional parametric diffusion process $dX_t = b(X_t, \alpha)\,dt + \sigma(X_t, \beta)$, $\alpha \in R^p$, $\beta \in R^q$, $p, q > = 1$, from discrete observations $\{X_{t_i}, i = 0, \ldots, n\}$ with $t_i = i\Delta_n$, $i = 0, 1, \ldots, n$, under the asymptotic scheme $\Delta_n \to 0$, $n\Delta_n \to \infty$ and $n\Delta_n^2 \to 0$. The class of ϕ-divergences is wide and includes several special members like Kullback–Leibler, Rényi, power and α-divergences. We derive the asymptotic distribution of the test statistics based on the estimated ϕ-divergences. The asymptotic distribution depends on the regularity of the function ϕ and in general it differs from the standard χ^2 distribution as in the i.i.d. case. Numerical analysis is used to show the small sample properties of the test statistics in terms of estimated level and power of the test.

1. Introduction

We consider the problem of parametric testing using ϕ-divergences. Let X be a random variable and $\{p(x, \theta), \theta \in \Theta \subset \mathbb{R}^m\}$, $m \geq 1$, a family of probability densities on the same measurable space. Let θ and θ_0 be two points in the interior of Θ and define the ϕ- divergence as follows:

$$D_\phi(\theta, \theta_0) = E_{\theta_0} \phi\left(\frac{p(X, \theta)}{p(X, \theta_0)}\right) = \int \phi\left(\frac{p(x, \theta)}{p(x, \theta_0)}\right) p(x, \theta_0)\,dx. \tag{1.1}$$

The function $\phi : [0, \infty) \to \mathbb{R}$, is assumed to be measurable, twice continuously differentiable, such that $\phi(1) = 0$ and (1.1) is well defined.

Examples of divergences of the form $D_\alpha(\theta, \theta_0) = D_{\phi_\alpha}(\theta, \theta_0)$ are the α-divergences, defined by means of the following function:

$$\phi_\alpha(x) = \frac{4(1 - x^{(1+\alpha)/2})}{1 - \alpha^2}, \quad -1 < \alpha < 1.$$

Note that $D_\alpha(\theta_0, \theta) = D_{-\alpha}(\theta, \theta_0)$. The class of α-divergences has been widely studied in statistics (see, e.g., Csiszár, 1967; Amari, 1985) and it is a family of divergences which includes several members of particular interest. For example, in the limit as $\alpha \to -1$, one obtains the well-known Kullback–Leibler distance

$$D_{-1}(\theta, \theta_0) = -E_{\theta_0} \log\left(\frac{p(X, \theta)}{p(X, \theta_0)}\right)$$

---

* Corresponding author.
  E-mail addresses: alessandro.degregorio@uniroma1.it (A. De Gregorio), stefano.iacus@unimi.it (S.M. Iacus).

while for $\alpha = 0$, the Hellinger distance (see, e.g., Beran, 1977; Simpson, 1989) emerges. As noticed in Chandra and Taniguchi (2006), the α- divergence is also equivalent to Rényi's (1961) divergence defined, for $\alpha \in (0, 1)$, as

$$R_\alpha(\theta, \theta_0) = \frac{1}{1-\alpha} \log E_{\theta_0} \left(\frac{p(X, \theta)}{p(X, \theta_0)} \right)^\alpha$$

from which is easy to see that in the limit as $\alpha \to 1$, R_α reduces to the Kullback–Leibler divergence. The transformation $\psi(R_\alpha) = (\exp\{(\alpha-1)R_\alpha - 1\}/(1-\alpha)$ returns the power-divergence studied in Cressie and Read (1984). Power-divergences can be obtained choosing $\phi(x) = \phi_\lambda(x)$ in (1.1), with

$$\phi_\lambda(x) = \frac{x^{\lambda+1} - \lambda(x-1) - x}{\lambda(\lambda+1)}, \quad \lambda \in \mathbb{R} - \{0, -1\}. \tag{1.2}$$

Liese and Vajda (1987) provided extensive study of a modified version of R_α and Morales et al. (1997) considered test statistics by means of ϕ-divergences with convex $\phi(\cdot)$ for independent and identically distributed (i.i.d) observations. A complete source of references about estimation and testing via ϕ-divergences can be found in Pardo (2006). In this paper we focus our attention on the ϕ-divergences for one-dimensional ergodic diffusion process $X = X_t, t \in [0, T]$, solution of the following stochastic differential equation:

$$dX_t = b(\alpha, X_t)\, dt + \sigma(\beta, X_t)\, dW_t, \quad X_0 = x_0, \tag{1.3}$$

with $\theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta$, $\Theta_\alpha \subset \mathbb{R}^p$ and $\Theta_\beta \subset \mathbb{R}^q$, $p, q \geq 1$. We denote the discrete time observations from X by $\mathbf{X}_n = \{X_{t_i}\}_{0 \leq i \leq n}, t_i = i\Delta_n, i = 0, 1, 2, \ldots, n$, where Δ_n is the length of the steps. The asymptotics is $\Delta_n \to 0, n\Delta_n \to \infty$ and $n\Delta_n^2 \to 0$ as $n \to \infty$. We study the asymptotic properties of the estimated ϕ- divergence for discretely observed diffusion processes defined as

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) = \phi\left(\frac{f_n(\mathbf{X}_n, \tilde{\theta}_n(\mathbf{X}_n))}{f_n(\mathbf{X}_n, \theta_0)} \right)$$

where $\theta_0 \in \Theta$, $f_n(\cdot, \cdot)$ is the approximated likelihood function proposed by Dacunha-Castelle and Florens-Zmirou (1986) and $\tilde{\theta}_n(\mathbf{X}_n)$ is any consistent and asymptotically normal estimator of θ. The hypotheses testing problem considered in this paper is the following:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

Our statistical procedure rejects the null hypothesis if $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) > c$, for some critical threshold c. The threshold c is determined using the asymptotic distribution of the test statistics $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0)$. We prove that $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0)$ converges in distribution to some function of the χ^2_{p+q} random variable. This result differs from the case of i.i.d. setting.

Up to our knowledge the only result concerning the use of divergences for discretely observed diffusion process is due to Rivas et al. (2005) where they consider the model of Brownian motion with drift $dX_t = a\, dt + b\, dW_t$ where a and b are two scalars. In that case, the exact likelihood of the observations is available in explicit form and it is the Gaussian law. Conversely, in the general setup of this paper, the likelihood of the process in (1.3) is known only for three particular stochastic differential equations, namely the Ornstein–Uhlenbeck diffusion, the geometric Brownian motion and the Cox–Ingersoll–Ross model. In all other cases, the likelihood has to be approximated. We choose the approximation due to Dacunha-Castelle and Florens-Zmirou (1986) and, to derive a proper estimator, we use the local Gaussian approximation proposed by Yoshida (1992) although our result holds for any consistent and asymptotically Gaussian estimator. This approach has been suggested by the work on Akaike information criteria for diffusion processes by Uchida and Yoshida (2005).

For continuous time observations from diffusion processes, Vajda (1990) considered the model $dX(t) = -b(t)X_t\, dt + \sigma(t)\, dW_t$; Küchler and Sørensen (1997) and Morales et al. (2004) contain several results on the likelihood ratio test statistics and Rényi statistics for exponential families. Explicit derivations of the Rényi information on the invariant law of ergodic diffusion processes have been presented in De Gregorio and Iacus (2009). For small diffusion processes, with continuous time observations, information criteria have been derived in Uchida and Yoshida (2004) using Malliavin calculus.

The problem of testing statistical hypotheses for general diffusion processes is still a developing stream of research. Kutoyants (2004) and Dachian and Kutoyants (2008) considered the problem of testing statistical hypotheses for ergodic diffusion models in continuous time; Kutoyants (1994) and Iacus and Kutoyants (2001) dealt with, respectively, parametric and semiparametric hypotheses testing for small diffusion processes; Negri and Nishiyama (2009, 2010) propose a non-parametric test based on score marked empirical process for both continuous and discrete time observation from small diffusion processes. The same test statistics has been studied for the ergodic case in Masuda et al. (2008). Lee and Wee (2008) considered the parametric version of the same test statistics for a simplified model. Aït-Sahalia (1996), Giet and Lubrano (2008) and Chen et al. (2008) introduced tests based on several distances between parametric and nonparametric estimation of the invariant density of discretely observed ergodic diffusion processes. The present paper complements the above references.

The paper is organized as follows. Section 2 introduces notation and regularity assumptions. Section 3 states the main result. Section 4 contains numerical experiments to study the small sample performance of the proposed test statistics in terms of empirical level and empirical power under some alternatives. The proofs are contained in Appendix A.

2. Assumptions on diffusion model

We consider a one-dimensional diffusion processes $X = X_t, t \in [0, T]$, solution to

$$dX_t = b(\alpha, X_t)\, dt + \sigma(\beta, X_t)\, dW_t, \quad X_0 = x_0, \tag{2.1}$$

where W_t is a Brownian motion. Let $\theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta$, where Θ_α and Θ_β are, respectively, compact convex subset of $\mathbb{R}^p$ and $\mathbb{R}^q$, $p, q \geq 1$. We assume that the drift function $b : \Theta_\alpha \times \mathbb{R} \to \mathbb{R}$ and the diffusion coefficient $\sigma : \Theta_\beta \times \mathbb{R} \to \mathbb{R}^+$ are known apart from the parameters α and β. Furthermore the process X_t is supposed ergodic for every θ with invariant law μ_θ, and observed at discrete times $t_i = i\Delta_n$, $i = 0, 1, 2, \ldots, n$, where Δ_n is the length of the steps. We denote the observations by $\mathbf{X}_n = \{X_{t_i}\}_{0 \leqslant i \leqslant n}$. The asymptotic is $\Delta_n \to 0$, $n\Delta_n \to \infty$ and $n\Delta_n^2 \to 0$ as $n \to \infty$.

In the definition of the ϕ- divergence (1.1) the likelihood function of the process is needed, but as noted in the introduction, it is usually not know. There are several ways to approximate the likelihood of a discretely observed diffusion process (for a review see, e.g., Iacus, 2008, Chapter 3). In this paper, we use the approximation proposed by Dacunha-Castelle and Florens-Zmirou (1986) although the result holds true (by changing some of the regularity conditions) for other approximations, like, e.g. the one based on Hermite polynomial expansion by Aït-Sahalia (2002) and even the local Gaussian approximation (see Yoshida, 1992; Kessler, 1997). To write it in explicit way, we use the same setup as in Uchida and Yoshida (2005). The following set of assumptions ensure the good behaviour of the approximated likelihood and the existence of a weak solution of (2.1)

Assumption 2.1 (*Regularity on the process*).

(i) There exists a positive constant C such that
$$|b(\alpha_0, x) - b(\alpha_0, y)| + |\sigma(\beta_0, x) - \sigma(\beta_0, y)| \leq C|x - y|.$$
(ii) $\inf_{\beta, x} \sigma^2(\beta, x) > 0$.
(iii) The process X is ergodic for every θ with invariant probability measure μ_θ.
(iv) For all $m \geq 0$ and for all θ, $\sup_t E|X_t|^m < \infty$.
(v) For every θ, the coefficients $b(\alpha, x)$ and $\sigma(\beta, x)$ are five times differentiable with respect to x and the derivatives are bounded by a polynomial function in x, uniformly in θ.
(vi) The coefficients $b(\alpha, x)$ and $\sigma(\beta, x)$ and all their partial derivatives respect to x up to order 2 are three times differentiable with respect to θ for all x in the state space. All derivatives with respect to θ are bounded by a polynomial function in x, uniformly in θ.

In order to define the approximated likelihood, we need to introduce the following functions:

$$s(x, \beta) = \int_0^x \frac{du}{\sigma(\beta, u)}, \quad B(x, \theta) = \frac{b(\alpha, x)}{\sigma(\beta, x)} - \frac{1}{2}\frac{\partial}{\partial x}\sigma(\beta, x),$$

$$\tilde{B}(x, \theta) = B\left(\frac{1}{s(\beta, x)}, \theta\right), \quad \tilde{h}(x, \theta) = \frac{\partial^2}{\partial x^2}\tilde{B}(x, \theta) + \frac{\partial}{\partial x}\tilde{B}(x, \theta).$$

Assumption 2.2 (*Regularity for the approximation, see Uchida and Yoshida, 2005*). For $i = 0, 1, 2, 3$ we have

(i) $(\partial^i/\partial\theta^i)\tilde{h}(x, \theta) = O(|x|^2)$ as $x \to \infty$.
(ii) $\inf_x (\partial^i/\partial\theta^i)\tilde{h}(x, \theta) > -\infty$.
(iii) $\sup_\theta \sup_x |(\partial^i/\partial\theta^i)(\partial^5/\partial x^5)\tilde{h}(x, \theta)| \leq M < \infty$.
(iv) There exists $\gamma > 0$ such that for every θ and $j = 1, \ldots, 4$, $|(\partial^i/\partial\theta^i)(\partial^j/\partial\theta^j)\tilde{B}(x, \theta)| = O(|\tilde{B}(x, \theta)|^\gamma)$ as $|x| \to \infty$.

Assumption 2.3 (*Identifiability*). If the coefficients $b(\alpha, x) = b(\alpha_0, x)$ and $\sigma^2(\beta, x) = \sigma^2(\beta_0, x)$ for all x (μ_{θ_0}—almost surely), then $\alpha = \alpha_0$ and $\beta = \beta_0$.

Under Assumptions 2.1 and 2.2 Dacunha-Castelle and Florens-Zmirou (1986) introduced the following approximation of the transition density f of the process X from y to x at lag t

$$f(t, x, y, \theta) = \frac{1}{\sqrt{2\pi t}\sigma(\beta, y)}\exp\left\{-\frac{S^2(x, y, \beta)}{2t} + H(x, y, \theta) + t\tilde{g}(x, y, \theta)\right\} \tag{2.2}$$

and its logarithm becomes

$$l(t, x, y, \theta) = -\frac{1}{2}\log(2\pi t) - \log\sigma(\beta, y) - \frac{S^2(x, y, \beta)}{2t} + H(x, y, \theta) + t\tilde{g}(x, y, \theta),$$

where

$$S(x,y,\beta) = \int_x^y \frac{du}{\sigma(\beta,u)}, \quad H(x,y,\theta) = \int_x^y \left\{ \frac{b(\alpha,u)}{\sigma^2(\beta,u)} - \frac{1}{2} \frac{\frac{\partial}{\partial x}\sigma(\beta,u)}{\sigma(\beta,u)} \right\} du,$$

$$\tilde{g}(x,y,\theta) = -\frac{1}{2} \left\{ C(x,\theta) + C(y,\theta) + \frac{1}{3} B(x,\theta) B(y,\theta) \right\},$$

$$C(x,\theta) = \frac{1}{3} B^2(x,\theta) + \frac{1}{2} \sigma(\beta,x) \frac{\partial}{\partial x} B(x,\theta).$$

The approximated likelihood and log-likelihood functions of the observations $\mathbf{X}_n$ become, respectively,

$$f_n(\mathbf{X}_n, \theta) = \prod_{i=1}^{n} f(\Delta_n, X_{t_i}, X_{t_{i-1}}, \theta) \tag{2.3}$$

and

$$l_n(\mathbf{X}_n, \theta) = \sum_{i=1}^{n} l(\Delta_n, X_{t_i}, X_{t_{i-1}}, \theta).$$

3. Construction of the test statistics and results

Clearly, the quantity $D_\phi(\theta, \theta_0)$ measures the discrepancy between two parametric models and represents an ideal candidate to construct a test statistics. Let $\tilde{\theta}_n(\mathbf{X}_n)$ be any consistent estimator of θ_0 and such that

$$\Gamma^{-1/2}(\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1}), \tag{3.1}$$

where $\mathcal{I}(\theta_0)$ is the positive definite and invertible Fisher information matrix at θ_0 equal to

$$\mathcal{I}(\theta_0) = \begin{pmatrix} (\mathcal{I}_b^{kj}(\theta_0))_{k,j=1,\dots,p} & 0 \\ 0 & (\mathcal{I}_\sigma^{kj}(\theta_0))_{k,j=1,\dots,q} \end{pmatrix},$$

where

$$\mathcal{I}_b^{kj}(\theta_0) = \int \frac{1}{\sigma^2(\beta_0,x)} \frac{\partial b(\alpha_0,x)}{\partial \alpha_k} \frac{\partial b(\alpha_0,x)}{\partial \alpha_j} \mu_{\theta_0}(dx),$$

$$\mathcal{I}_\sigma^{kj}(\theta_0) = 2 \int \frac{1}{\sigma^2(\beta_0,x)} \frac{\partial \sigma(\beta_0,x)}{\partial \beta_k} \frac{\partial \sigma(\beta_0,x)}{\partial \beta_j} \mu_{\theta_0}(dx).$$

We denote by Γ the $(p+q) \times (p+q)$ matrix

$$\Gamma = \begin{pmatrix} \frac{1}{n\Delta_n} I_p & 0 \\ 0 & \frac{1}{n} I_q \end{pmatrix},$$

where I_p is the $p \times p$ identity matrix.

To construct the test statistics by means of an estimate of the ϕ- divergence in (1.1), we make use of the approximated likelihood $f_n(\mathbf{X}_n, \theta)$ and replace θ by any consistent and asymptotically Gaussian estimator $\tilde{\theta}_n(\mathbf{X}_n)$. Having only one single observation of $\mathbf{X}_n$, i.e. only one observed trajectory, we estimate (1.1) with

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) = \phi\left(\frac{f_n(\mathbf{X}_n, \tilde{\theta}_n(\mathbf{X}_n))}{f_n(\mathbf{X}_n, \theta_0)} \right). \tag{3.2}$$

Conversely to the i.i.d. case, there is no expected value in (3.2). We will discuss this point after the presentation of Theorem 3.1. We consider the following hypotheses testing problem:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

Our statistical procedure rejects the null hypothesis if $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) > c$, for some critical threshold c. Next theorem gives the asymptotic distribution of $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0)$ from which it is possible to determine the threshold c.

Consider the divergence defined in (3.2) and let $\phi : [0, \infty) \to \mathbb{R}$ be measurable, twice continuously differentiable and such that $\phi(1) = 0$. We denote by $C_\phi = \phi'(1)$ and $K_\phi = \phi''(1)$ the first two derivatives at point 1. Assume that at least one among C_ϕ and K_ϕ is not zero.

Theorem 3.1. *Under* $H_0 : \theta = \theta_0$, *Assumptions 2.1–2.3, convergence* (3.1), *we have that*

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \xrightarrow{d} \tfrac{1}{2}(C_\phi \xi_{p+q} + (C_\phi + K_\phi)\xi_{p+q}^2), \tag{3.3}$$

where $\xi_{p+q} \sim \chi_{p+q}^2$ *is the Chi-squared random variable with* $p+q$ *degrees of freedom.*

Remark 3.1. The proof of Theorem 3.1 is based on Taylor expansion up to second order like in the i.i.d. case. But if one uses first order approximation, it is easy to obtain that $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \xrightarrow{d} C_\phi \chi_{p+q}^2$ when $C_\phi \neq 0$.

Remark 3.2. If we consider the limit as $\alpha \to -1$ for $\phi_\alpha(x)$ of the α-divergences, i.e. we consider the Kullback–Leibler divergence, we have

$$\phi_{-1}(x) := \lim_{\alpha \to -1} \phi_\alpha(x) = -\log(x)$$

for which $C_\phi = -1$ and $K_\phi = 1$. In that case, (3.3) reduces to the standard result for the likelihood ratio test statistics. For the power-divergences, with ϕ as in (1.2), it is easy to verify that $C_\phi = \phi'(1) = 0$ and the asymptotic distribution differs from the usual i.i.d. setting, i.e. $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \xrightarrow{d} \tfrac{1}{2} K_\phi Z_{p+q}$.

The convergence in Theorem 3.1 may appear somewhat strange if one thinks about the usual results on ϕ-divergences for i.i.d. observations. The main difference in diffusion models, is that our estimate of the divergence has not the usual form of an expected value, i.e. it estimates the expected value with one observation only. This is why, in the i.i.d case, the first term in the Taylor expansion of D_ϕ vanishes being the expected value of the score function, while in our case it remains only the score function which, as usual, converges to a Gaussian random variable. For the same reason, in the second term of the Taylor expansion, in the i.i.d. case the expected value of the second order derivative appears which converges to the Fisher information and, in our case, we have no expected value, hence the convergence to the square of the χ^2 emerges.

If one wants to emulate the standard results for the i.i.d. case, it is still possible to work on the invariant density of the diffusion process. In that case, the ϕ- divergence takes the usual form of the i.i.d. case because the invariant density have an explicit form. Indeed, let

$$s(x, \theta) = \exp\left\{-2 \int_{\tilde{x}}^x \frac{b(\theta, y)}{\sigma^2(\theta, y)} \, dy\right\}, \quad m(x, \theta) = \frac{1}{\sigma^2(\theta, x) s(x, \theta)}$$

be the scale and speed functions of the diffusion, with $\tilde{x}$ some value in the state space of the diffusion process. Let $M = \int m(x, \theta) \, dx < \infty$, then $\pi(x, \theta) = m(x, \theta)/M$ is the invariant density of the diffusion process. In this case, it is possible to define the ϕ- divergence as

$$D_\phi(\tilde{\theta}_n, \theta_0) = \int \phi\left(\frac{\pi(x, \tilde{\theta}_n)}{\pi(x, \theta_0)}\right) \pi(x, \theta_0) \, dx$$

and the standard results follows.

Remark 3.3. In our application, to derive and estimator, we consider further the local Gaussian approximation of the same transition density (see, Yoshida, 1992)

$$g_n(\mathbf{X}_n, \theta) = \sum_{i=1}^n g_n(\Delta_n, X_{t_{i-1}}, X_{t_i}, \theta), \tag{3.4}$$

where

$$g(t, y, x, \theta) = -\frac{1}{2}\log(2\pi t) - \log \sigma(\beta, x) - \frac{[y - x - tb(\alpha, x)]^2}{2t\sigma^2(\beta, x)}.$$

The approximate maximum likelihood estimator $\hat{\theta}_n(\mathbf{X}_n)$ based on (3.4) is then defined as

$$\hat{\theta}_n(\mathbf{X}_n) = \arg\sup_\theta g_n(\mathbf{X}_n, \theta). \tag{3.5}$$

Under the condition $n\Delta_n^2 \to 0$ (see Theorem 1 in Kessler, 1997) the estimator $\hat{\theta}_n(\mathbf{X}_n)$ in (3.5) satisfies (3.1). Hence, the result of Theorem 3.1 applies for $\tilde{\theta}_n(\mathbf{X}_n) = \hat{\theta}_n(\mathbf{X}_n)$.

Remark 3.4. In Theorem 3.1 there is no need to impose $C_\phi = 0$ and $K_\phi = 1$ as, e.g. in Morales et al. (1997). Of course, in our case the constants C_ϕ and K_ϕ enter in the asymptotic distribution of the test statistics. The convergence result is also interesting because, contrary to the i.i.d case, the rate of convergence of the estimators of θ for the drift and diffusion coefficients are different and are, respectively, equal to $\sqrt{n\Delta_n}$ and $\sqrt{n}$.

Remark 3.5. Assumptions 2.2 may appear a bit complicated to check, but the result in Theorem 3.1 holds even if one uses the local Gaussian approximation in Yoshida (1992) or the Hermite polynomial expansion in Aït-Sahalia (2002) instead of (2.3). In that case one just need to replace Assumption 2.2 with the corresponding regularity conditions. The proof of Theorem 3.1 remains unchanged.

4. Numerical analysis

Although asymptotic properties have been obtained, what really matters in application is the behaviour of the test statistics under fine sample setup. We study the empirical performance of the test for small samples in terms of level of the test and power under some alternatives. In the analysis we consider the estimator (3.5) and compare the power divergences

$$\mathbb{D}_\lambda(\hat{\theta}_n(\mathbf{X}_n), \theta_0) = \phi_\lambda\left(\frac{f_n(\mathbf{X}_n, \hat{\theta}_n(\mathbf{X}_n))}{f_n(\mathbf{X}_n, \theta_0)}\right),$$

with $\phi_\lambda(x) = (x^{\lambda+1} - x - \lambda(x-1))/(\lambda(\lambda+1))$, $C_\lambda = 0$, $K_\lambda = 1$ and $\lambda \in \{-0.99, -1.20, -1.50, -1.75, -2.00, -2.50\}$, and the generalized likelihood ratio test

$$\mathbb{D}_{\log}(\hat{\theta}_n(\mathbf{X}_n), \theta_0) = -\log\left(\frac{f_n(\mathbf{X}_n, \hat{\theta}_n(\mathbf{X}_n))}{f_n(\mathbf{X}_n, \theta_0)}\right).$$

For $\mathbb{D}_{\log}$ and $\mathbb{D}_\lambda$, the threshold of the rejection region of the test are calculated, respectively, using the true quantiles of the χ^2_{p+q} random variable and using formula (3.3) and the empirical quantiles of (3.3) of 100,000 simulations of the random variable χ^2_{p+q}.

We evaluate the empirical level of the test calculated as the number of times the test rejects the null hypothesis under the true model, i.e.

$$\hat{\alpha}_n = \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}_{\{\mathbb{D}_\phi > c_\alpha\}},$$

where $\mathbf{1}_A$ is the indicator function of set A, $M = 10,000$ is the number of simulations and c_α is the $(1-\alpha)\%$ quantile of the proper distribution. Similarly we calculate the power of the test under alternative models as

$$\hat{\beta}_n = \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}_{\{\mathbb{D}_\phi > c_\alpha\}}.$$

The simulation setup is the same as the ones proposed in Pritsker (1998) and Chen et al. (2008). The parameters of the chosen models under the null hypotheses H_0 and the alternatives H_1, have been set according to these authors. In particular VAS_0 corresponds to the model estimated by Aït-Sahalia (1996) for real interest rates data. In our experiments we consider then the following two families of stochastic processes:

The Vasicek (VAS) model: $dX_t = \kappa(\alpha - X_t)\,dt + \sigma\,dW_t$, where, in finance, σ is interpreted as volatility, α is the long-run equilibrium value of the process and κ is the speed of reversion. Let

$$H_0 : \theta = \theta_0 = (\kappa_0, \alpha_0, \sigma_0^2) = (0.85837, 0.089102, 0.0021854) \quad \text{model } VAS_0$$

we consider two different alternative hypotheses H_1

$$H_1 : \theta = \theta_1 = (4 \cdot \kappa_0, \alpha_0, 4 \cdot \sigma_0^2) \quad \text{model } VAS_1$$

and

$$H_1 : \theta = \theta_1 = (\tfrac{1}{4}\kappa_0, \alpha_0, \tfrac{1}{4}\cdot\sigma_0^2) \quad \text{model } VAS_2.$$

The interesting facts are that VAS_0, VAS_1 and VAS_2 have all the same stationary distributions $N(\alpha_0, \sigma_0^2/(2\kappa_0))$, a Gaussian transition density

$$N\left(\alpha_0 + (x_0 - \alpha_0)e^{-\kappa t}, \frac{\sigma_0^2(1 - e^{-2\kappa t})}{2\kappa_0}\right)$$

and covariance function given by

$$\text{Cov}(X_s, X_t) = \frac{\sigma_0^2}{2\kappa_0}e^{-\kappa(s+t)}(e^{-2\kappa\min(s,t)-1}).$$

Moreover, both show a strong dependency of the covariance as a function of κ, which makes this model interesting in comparison with the i.i.d. setting.

The Cox–Ingersoll–Ross (CIR) model: $dX_t = \kappa(\alpha - X_t)\,dt + \sigma\sqrt{X_t}\,dW_t$. Let $(\kappa_0, \alpha_0, \sigma_0^2) = (0.89218, 0.09045, 0.032742)$, we consider different values of the parameters under H_0 and H_1 like in the Vasicek case, i.e.

Model	$\theta = (\kappa, \alpha, \sigma^2)$
CIR_0	$\theta_0 = (\kappa_0, \alpha_0, \sigma_0^2)$
CIR_1	$\theta_1 = (\tfrac{1}{2}\cdot\kappa_0, \alpha_0, \tfrac{1}{2}\cdot\sigma_0^2)$
CIR_2	$\theta_1 = (\tfrac{1}{4}\cdot\kappa_0, \alpha_0, \tfrac{1}{4}\cdot\sigma_0^2)$

The CIR model has a transition density of χ^2- type, hence local Gaussian approximation is less likely to hold for non negligible values of Δ_n, so this may impact the performance of the estimator $\hat{\theta}_n$. Notice that the general CIR model does not satisfy Assumption 2.1(ii), but for our particular choice of the parameter set, the process is strictly positive for all x, so the approximation holds.

We study the empirical level and the empirical power of the different test statistics for different values of $\Delta_n \in \{0.1, 0.001\}$ and $n \in \{50, 100, 500\}$. For each trajectory we simulate 1000 observations and extract only that last n observations. Disregarding the first part of the trajectory ensures that the process is in the stationary state.

The results of these simulations are reported in Tables 1 and 2. We point out that in the tables in the column "model (α, n)" the α corresponds to the true level of the test used to calculate c_α. The column $\mathbb{D}_{\log}$ reports the empirical level and power for the generalized likelihood test which is our benchmark and the remaining columns the same quantities for the power-divergences $\mathbb{D}_\lambda$ for different values of λ.

Summary of the analysis for the Vasicek model: It emerges that for $\lambda = -0.99$, the power divergence cannot identify as wrong model VAS_1 for small sample size $n = 50$ and $\Delta_n = 0.001$ (Table 1, row 2), although this is not the case for the other values of λ and the likelihood ratio test.

In general power divergences for λ in $\{-0.99, -1.20, -1.50, -1.75, -2.00\}$ have always very small empirical level and high empirical power under the selected alternatives.

The power divergences are, on average, better than the likelihood ratio test in terms of both empirical level $\hat{\alpha}$ and power $\hat{\beta}$ under the selected alternatives.

Summary of the analysis for the CIR model: As seen from Table 2, the same average considerations apply to the case of CIR model. The difference is that, for small sample size, all test statistics have low power under the alternative CIR_1 while CIR_2 does not present particular problems.

5. Conclusions

It seems that, as in the i.i.d. case, also for discretely observed diffusion processes the ϕ-divergences may compete or improve the performance of the standard likelihood ratio statistics. In particular, the power divergences are in general quite good in terms of estimated level and power of the test even for moderate sample sizes (e.g. $n \geq 100$ in our simulations).

The package `sde` for the `R` statistical environment (R Development Core Team, 2009) and freely available at http://cran.R-Project.org contains the function `sdeDiv` which implements the ϕ- divergence test statistics.

Acknowledgements

Appendix A. Proofs

The following important lemmas are useful to prove Theorem 3.1.

Lemma A.1 (*Kessler, 1997*). Under Assumptions 2.1–2.3, the following hold true:

$$\Gamma^{1/2}\nabla_\theta g_n(\mathbf{X}_n, \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)). \tag{A.1}$$

Lemma A.2 (*Uchida and Yoshida, 2005*). Under the Assumptions 2.1–2.3, the following hold true:

$$\Gamma^{1/2}\nabla_\theta l_n(\mathbf{X}_n, \theta_0) = \Gamma^{1/2}\nabla_\theta g_n(\mathbf{X}_n, \theta_0) + o_p(1). \tag{A.2}$$

Lemma A.3 (*Uchida and Yoshida, 2005*). Under Assumptions 2.1–2.3, the following hold true:

$$\Gamma^{1/2}\nabla_\theta^2 l_n(\mathbf{X}_n, \theta_0)\Gamma^{1/2} \xrightarrow{P} -\mathcal{I}(\theta_0). \tag{A.3}$$

Proof of Theorem 3.1. We start by applying *delta* method. We denote the gradient vector by $\nabla_\theta = [\partial/\partial\theta_i]$, $i = 1, \ldots, p+q$ and similarly the Hessian matrix by $\nabla_\theta^2 = [\partial^2/\partial\theta_i\partial\theta_j]$, $i,j = 1, \ldots, p+q$. We have

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) = \mathbb{D}_\phi(\theta_0, \theta_0) + [\Gamma^{1/2}\nabla_\theta\mathbb{D}_\phi(\theta_0, \theta_0)]^T\Gamma^{-1/2}(\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) + \frac{1}{2}[\Gamma^{-1/2}(\tilde{\theta}_n(\mathbf{X}_n) - \theta_0)]^T\Gamma^{1/2}\nabla_\theta^2\mathbb{D}_\phi(\theta_0, \theta_0)\Gamma^{1/2}$$

$$\times \Gamma^{-1/2}(\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) + o_p(1). \tag{A.4}$$

Notice that $\mathbb{D}_\phi(\theta_0, \theta_0) = 0$ and, for $k = 1, \ldots, p+q$,

$$\frac{\partial}{\partial\theta_k}\left[\phi\left(\frac{f_n(\cdot, \theta)}{f_n(\cdot, \theta_0)}\right)\right] = \frac{1}{f_n(\cdot, \theta_0)}\phi'\left(\frac{f_n(\cdot, \theta)}{f_n(\cdot, \theta_0)}\right)\frac{\partial f_n(\cdot, \theta)}{\partial\theta_k}$$

Table 1
Numbers represent probability of rejection under the true generating model, with c_α calculated under H_0.

Model (α, n)	$\mathbb{D}_{\log}$	$\lambda = -0.99$	$\lambda = -1.20$	$\lambda = -1.50$	$\lambda = -1.75$	$\lambda = -2.00$	$\lambda = -2.50$
VAS_0 (0.01, 50)	0.01	0.00	0.00	0.00	0.01	0.02	0.04
VAS_1 (0.01, 50)	1.00	0.00	0.99	1.00	1.00	1.00	1.00
VAS_2 (0.01, 50)	1.00	0.40	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.05, 50)	0.04	0.00	0.00	0.00	0.01	0.03	0.06
VAS_1 (0.05, 50)	1.00	0.67	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00	
VAS_0 (0.01, 100)	0.01	0.00	0.00	0.00	0.01	0.02	0.04
VAS_1 (0.01, 100)	1.00	0.23	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.01, 100)	1.00	0.88	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.05, 100)	0.04	0.00	0.00	0.00	0.01	0.03	0.06
VAS_1 (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.01, 500)	0.02	0.00	0.00	0.00	0.01	0.03	0.08
VAS_1 (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.05, 500)	0.07	0.00	0.00	0.01	0.03	0.06	0.12
VAS_1 (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.01, 50)	0.01	0.00	0.00	0.00	0.01	0.02	0.05
VAS_1 (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.05, 50)	0.05	0.00	0.00	0.00	0.02	0.03	0.09
VAS_1 (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.01, 100)	0.01	0.00	0.00	0.00	0.00	0.01	0.05
VAS_1 (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.05, 100)	0.04	0.00	0.00	0.00	0.01	0.03	0.08
VAS_1 (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.01, 500)	0.00	0.00	0.00	0.00	0.00	0.01	0.02
VAS_1 (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_0 (0.05, 500)	0.02	0.00	0.00	0.00	0.01	0.01	0.04
VAS_1 (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAS_2 (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Therefore, the values are $\hat{\alpha}$ under model "0" and $\hat{\beta}$ otherwise. Estimates calculated on 10,000 experiments. Estimated power-divergences for $\Delta_n = 0.001$ (upper table) and $\Delta_n = 0.1$ (bottom table).

and, for $k, j = 1, \ldots, p+q$,

$$\frac{\partial^2}{\partial \theta_k \partial \theta_j} \left[\phi\left(\frac{f_n(\cdot, \theta)}{f_n(\cdot, \theta_0)}\right) \right] = \frac{1}{f_n^2(\cdot, \theta_0)} \phi''\left(\frac{f_n(\cdot, \theta)}{f_n(\cdot, \theta_0)}\right) \frac{\partial f_n(\cdot, \theta)}{\partial \theta_k} \frac{\partial f_n(\cdot, \theta)}{\partial \theta_j} + \frac{1}{f_n(\cdot, \theta_0)} \phi'\left(\frac{f_n(\cdot, \theta)}{f_n(\cdot, \theta_0)}\right) \frac{\partial^2 f_n(\cdot, \theta)}{\partial \theta_k \partial \theta_j}.$$

Therefore

$$\nabla_\theta \mathbb{D}_\phi(\theta_0, \theta_0) = C_\phi \nabla_\theta l_n(\mathbf{X}_n, \theta)\big|_{\theta = \theta_0} = C_\phi \nabla_\theta l_n(\mathbf{X}_n, \theta_0) \tag{A.5}$$

Table 2
Numbers represent probability of rejection under the true model, with rejection region calculated under H_0.

Model (α, n)	$\mathbb{D}_{\log}$	$\lambda = -0.99$	$\lambda = -1.20$	$\lambda = -1.50$	$\lambda = -1.75$	$\lambda = -2.00$	$\lambda = -2.50$
CIR_0 (0.01, 50)	0.03	0.00	0.00	0.00	0.02	0.05	0.13
CIR_1 (0.01, 50)	0.63	0.00	0.01	0.28	0.55	0.71	0.87
CIR_2 (0.01, 50)	1.00	0.00	0.81	1.00	1.00	1.00	1.00
CIR_0 (0.05, 50)	0.12	0.00	0.00	0.01	0.04	0.09	0.19
CIR_1 (0.05, 50)	0.86	0.00	0.09	0.48	0.70	0.81	0.92
CIR_2 (0.05, 50)	1.00	0.00	0.97	1.00	1.00	1.00	1.00
CIR_0 (0.01, 100)	0.03	0.00	0.00	0.00	0.02	0.05	0.14
CIR_1 (0.01, 100)	0.97	0.00	0.21	0.83	0.95	0.98	0.99
CIR_2 (0.01, 100)	1.00	0.00	1.00	1.00	1.00	1.00	1.00
CIR_0 (0.05, 100)	0.12	0.00	0.00	0.01	0.05	0.09	0.20
CIR_1 (0.05, 100)	0.99	0.00	0.53	0.93	0.98	0.99	1.00
CIR_2 (0.05, 100)	1.00	0.24	1.00	1.00	1.00	1.00	1.00
CIR_0 (0.01, 500)	0.03	0.00	0.00	0.00	0.02	0.04	0.10
CIR_1 (0.01, 500)	1.00	0.00	1.00	1.00	1.00	1.00	1.00
CIR_2 (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CIR_0 (0.05, 500)	0.09	0.00	0.00	0.01	0.04	0.07	0.15
CIR_1 (0.05, 500)	1.00	0.95	1.00	1.00	1.00	1.00	1.00
CIR_2 (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CIR_0 (0.01, 50)	0.01	0.00	0.00	0.00	0.01	0.02	0.06
CIR_1 (0.01, 50)	0.80	0.00	0.06	0.52	0.75	0.86	0.94
CIR_2 (0.01, 50)	1.00	0.00	0.99	1.00	1.00	1.00	1.00
CIR_0 (0.05, 50)	0.05	0.00	0.00	0.00	0.02	0.04	0.09
CIR_1 (0.05, 50)	0.94	0.00	0.23	0.70	0.85	0.92	0.96
CIR_2 (0.05, 50)	1.00	0.06	1.00	1.00	1.00	1.00	1.00
CIR_0 (0.01, 100)	0.01	0.00	0.00	0.00	0.01	0.02	0.05
CIR_1 (0.01, 100)	0.99	0.00	0.56	0.96	0.99	1.00	1.00
CIR_2 (0.01, 100)	1.00	0.00	1.00	1.00	1.00	1.00	1.00
CIR_0 (0.05, 100)	0.04	0.00	0.00	0.00	0.02	0.03	0.08
CIR_1 (0.05, 100)	1.00	0.00	0.83	0.99	1.00	1.00	1.00
CIR_2 (0.05, 100)	1.00	0.97	1.00	1.00	1.00	1.00	1.00
CIR_0 (0.01, 500)	0.00	0.00	0.00	0.00	0.00	0.01	0.02
CIR_1 (0.01, 500)	1.00	0.00	1.00	1.00	1.00	1.00	1.00
CIR_2 (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CIR_0 (0.05, 500)	0.02	0.00	0.00	0.00	0.01	0.02	0.04
CIR_1 (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CIR_2 (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Therefore, the values are $\hat{\alpha}$ under model "0" and $\hat{\beta}$ otherwise. Estimates calculated on 10,000 experiments. Estimated power-divergences for $\Delta_n = 0.001$ (upper table) and $\Delta_n = 0.1$ (bottom table).

and

$$\nabla_\theta^2 \mathbb{D}_\phi(\theta_0, \theta_0) = K_\phi \nabla_\theta l_n(\mathbf{X}_n, \theta_0)[\nabla_\theta l_n(\mathbf{X}_n, \theta_0)]^T + C_\phi \frac{1}{f(\mathbf{X}_n, \theta_0)} \nabla_\theta^2 f(\mathbf{X}_n, \theta_0)$$

$$= (K_\phi + C_\phi) \nabla_\theta l_n(\mathbf{X}_n, \theta_0)[\nabla_\theta l_n(\mathbf{X}_n, \theta_0)]^T + C_\phi \nabla_\theta^2 l_n(\mathbf{X}_n, \theta_0). \tag{A.6}$$

Plugging in (A.4) the terms (A.5) and (A.6) and making use of Lemmas A.1–A.3 and Slutsky's theorem we get

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \xrightarrow{d} \tfrac{1}{2}[C_\phi \xi_{p+q} + (C_\phi + K_\phi)\xi_{p+q}^2],$$

where $\xi_{p+q} \sim \chi^2_{p+q}$. It is easy to verify that the density function of the r.v. ξ^2_{p+q} is equal to

$$f_{\xi^2_{p+q}}(z) = \frac{(1/2)^{(p+q)/2}}{\Gamma\left(\frac{p+q}{2}\right)} \sqrt{z}^{(p+q)/2-1} e^{-\sqrt{z}/2} \frac{1}{2\sqrt{z}}, \quad z > 0. \qquad \square \tag{A.7}$$

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Rev. Financ. Stud. 70 (2), 385–426.

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

Amari, S., 1985. Differential-Geometrical Methods in Statistics. In: Lecture Notes in Statistics, vol. 28. Springer, New York.

Beran, R.J., 1977. Minimum Hellinger estimates for parametric models. Ann. Stat. 5, 445–463.

Chandra, S.A., Taniguchi, M., 2006. Minimum α- divergence estimation for arch models. J. Time Ser. Anal. 27 (1), 19–39.

Chen, S.X., Gao, J., Cheng, Y.T., 2008. A test for model specification of diffusion processes. Ann. Statist. 36 (1), 167–198.

Cressie, N., Read, T.R.C., 1984. Multinomial goodness of fit tests. J. R. Stat. Soc. Ser. B Stat. Methodol. 46, 440–464.

Csiszár, I., 1967. On topological properties of *f*-divergences. Studia Sci. Math. Hungar. 2, 329–339.

Dachian, S., Kutoyants, Yu.A., 2008. On the goodness-of-t tests for some continuous time processes. In: Vonta, F., Nikulin, M., Limnios, N., Huber-Carol, C. (Eds.), Statistical Models and Methods for Biomedical and Technical Systems. Birkhuser, Boston, pp. 395–413.

Dacunha-Castelle, D., Florens-Zmirou, D., 1986. Estimation of the coefficients of a diffusion from discrete observations. Stochastics 19, 263–284.

De Gregorio, A., Iacus, S.M., 2009. Rényi information for ergodic diffusion processes. Inform. Sci. 179, 279–291.

Giet, L., Lubrano, M., 2008. A minimum Hellinger distance estimator for stochastic differential equations: an application to statistical inference for continuous time interest rate models. Comput. Statist. Data Anal. 52, 2945–2965.

Iacus, S.M., 2008. Simulation and Inference for Stochastic Differential Equations. Springer, New York.

Iacus, S.M., Kutoyants, Y., 2001. Semiparametric hypotheses testing for dynamical systems with small noise. Math. Methods Statist. 10 (1), 105–120.

Kessler, M., 1997. Estimation of an ergodic diffusion from discrete observations. Scand. J. Statist. 24, 211–229.

Kutoyants, Y., 1994. Identification of Dynamical Systems with Small Noise. Kluwer, Dordrecht.

Kutoyants, Y., 2004. Statistical Inference for Ergodic Diffusion Processes. Springer, London.

Küchler, U., Sørensen, M., 1997. Exponential Families of Stochastic Processes. Springer, New York.

Lee, S., Wee, I.-S., 2008. Residual empirical process for diffusion processes. J. Korean Math. Soc. 45 (3).

Liese, F., Vajda, I., 1987. Convex Statistical Distances. Tuebner, Leipzig.

Masuda, H., Negri, I., Nishiyama, Y., 2008. Goodness of fit test for ergodic diffusions by discrete time observations: an innovation martingale approach. Research Memorandum 1069, Inst. Statist. Math., Tokyo.

Morales, D., Pardo, L., Vajda, I., 1997. Some new statistics for testing hypotheses in parametric models. J. Multivariate Anal. 67, 137–168.

Morales, D., Pardo, L., Pardo, M.C., Vajda, I., 2004. Rényi statistics in directed families of exponential experiments. Statistics 38 (2), 133–147.

Negri, I., Nishiyama, Y., 2009. Goodness of fit test for ergodic diffusion processes. Ann. Inst. Statist. Math. 61, 919–928.

Negri, I., Nishiyama, Y., 2010. Goodness of fit test for small diffusions based on discrete observations. Ann. Inst. Stat. Math., to appear.

Pardo, L., 2006. Statistical Inference Based on Divergence Measures. Chapman & Hall/CRC, London.

Pritsker, M., 1998. Nonparametric density estimation and tests of continuous time interest rate models. Rev. Financ. Stud. 11, 449–487.

R Development Core Team, 2009. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL ⟨http://www.R-project.org⟩.

Rényi, A., 1961. On measures of entropy and information. In: Proceedings of the Fourth Berkeley Symposium on Probability and Mathematical Statistics, vol. 1, University of California, Berkeley, pp. 547–461.

Rivas, M.J., Santos, M.T., Morales, D., 2005. Rényi test statistics for partially observed diffusion processes. J. Statist. Plann. Inference 127, 91–102.

Simpson, D.G., 1989. Hellinger deviance tests: efficiency, breakdown points, and examples. J. Amer. Statist. Assoc. 84, 107–113.

Uchida, M., Yoshida, N., 2004. Information criteria for small diffusions via the theory of Malliavin–Watanabe. Stat. Inference Stoch. Process. 7, 35–67.

Uchida, M., Yoshida, N., 2005. AIC for ergodic diffusion processes from discrete observations, preprint MHF 2005-12, March 2005. Faculty of Mathematics, Kyushu University, Fukuoka, Japan.

Vajda, I., 1990. Rényi distances of some diffusion processes. In: Statulevičius, V. (Ed.), Probability Theory and Mathematical Statistics. Mokslas, Vilnius, pp. 529–534.

Yoshida, N., 1992. Estimation for diffusion processes from discrete observation. J. Multivar. Anal. 41 (2), 220–242.