

# Divergences Test Statistics for Discretely Observed Diffusion Processes

Alessandro De Gregorio<sup>a</sup>, Stefano M. Iacus<sup>b,\*</sup>

<sup>a</sup>*Department of Statistics, Probability and Applied Statistics  
P.le Aldo Moro 5, 00185 Rome-Italy, Email: alessandro.degregorio@uniroma1.it*

<sup>b</sup>*Department of Economics, Business and Statistics  
Via Conservatorio 7, 20124 Milan-Italy, Email: stefano.iacus@unimi.it*

---

## Abstract

In this paper we propose the use of  $\phi$ -divergences as test statistics to verify simple hypotheses about a one-dimensional parametric diffusion process  $dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t$ , from discrete observations  $\{X_{t_i}, i = 0, \dots, n\}$  with  $t_i = i\Delta_n$ ,  $i = 0, 1, \dots, n$ , under the asymptotic scheme  $\Delta_n \rightarrow 0$ ,  $n\Delta_n \rightarrow \infty$  and  $n\Delta_n^2 \rightarrow 0$ . The class of  $\phi$ -divergences is wide and includes several special members like Kullback-Leibler, Rényi, power and  $\alpha$ -divergences. We derive the asymptotic distribution of the test statistics based on  $\phi$ -divergences. The limiting law takes different forms depending on the regularity of  $\phi$ . These convergence differ from the classical results for independent and identically distributed random variables. Numerical analysis is used to show the small sample properties of the test statistics in terms of estimated level and power of the test.

*Key words:* diffusion processes, empirical level, hypotheses testing,  $\phi$ -divergences,  $\alpha$ -divergences

---

## 1. Introduction

We consider the problem of parametric testing using  $\phi$ -divergences. Let  $X$  be a r.v. and  $f(X, \theta)$  and  $g(X, \theta)$ ,  $\theta \in \Theta$  two families of probability densities on the same measurable space. The  $\phi$ -divergences are defined as  $D_\phi(f, g) = E_\theta \phi(f(X)/g(X))$ , where  $E_\theta$  is the expected value with respect to

---

\*Corresponding author

$P_\theta$ , the true law of the observations. Because we focus the attention on the use of divergences for hypotheses testing, we will use a simplified notation: let  $\theta$  and  $\theta_0$  two points in the interior of  $\Theta$  and define the divergence as

$$D_\phi(\theta, \theta_0) = E_{\theta_0} \phi \left( \frac{p(X, \theta)}{p(X, \theta_0)} \right) \quad (1.1)$$

In equation (1.1) the density  $\{p(X, \theta), \theta \in \Theta\}$  is a same family of probability densities and  $\phi(\cdot)$  is a function with the minimal property that  $\phi(1) = 0$ . Examples of divergences of the form  $D_\alpha(\theta, \theta_0) = D_{\phi_\alpha}(\theta, \theta_0)$  are the  $\alpha$ -divergences, defined by means of the following function

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

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

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

while as  $\alpha \rightarrow 0$ , the Hellinger distance (see, e.g., Beran, 1977, Simpson, 1989) emerges

$$D_0(\theta, \theta_0) = \frac{1}{2} E \left( \sqrt{p(X, \theta)} - \sqrt{p(X, \theta_0)} \right)^2$$

As noticed in Chandra and Taniguchi (2006), the  $\alpha$ -divergence is also equivalent to the Rényi's divergence (Rényi, 1961) 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  $\alpha \rightarrow 1$ ,  $R_\alpha$  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). Liese and Vajda (1987) provide extensive study of a modified version of  $R_\alpha$  and Morales *et al.* (1997) consider divergences with convex  $\phi(\cdot)$  for independent and identically distributed (i.i.d) observations; for example the

power-divergences  $D_{\phi_\lambda}(\theta, \theta_0)$  with

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

In this paper we focus our attention on the  $\phi$ -divergences  $D_\phi(\theta, \theta_0)$ , defined as in (1.1), for one-dimensional ergodic diffusion process  $\{X_t, t \in [0, T]\}$ , solution of the following stochastic differential equation

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

with  $\theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta$ ,  $\Theta_\alpha \subset \mathbb{R}^p$  and  $\Theta_\beta \subset \mathbb{R}^q$ . We indicate the discrete time observations from  $X$  as  $\mathbf{X}_n = \{X_{t_i}\}_{0 \leq i \leq n}$ ,  $t_i = i\Delta_n$ ,  $i = 0, 1, 2, \dots, n$ , where  $\Delta_n$  is the length of the steps. The asymptotics is  $\Delta_n \rightarrow 0$ ,  $n\Delta_n \rightarrow \infty$  and  $n\Delta_n^2 \rightarrow 0$  as  $n \rightarrow \infty$ . We study the properties of the estimated  $\phi$ -divergence  $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0)$ , 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  $f_n(\cdot, \cdot)$  is the approximated likelihood proposed by Dacunha-Castelle and Florens-Zmirou (1986) and  $\tilde{\theta}_n(\mathbf{X}_n)$  is any consistent, asymptotically normal and efficient estimator of  $\theta$ . We prove that, for  $\phi(\cdot)$  functions which satisfying three different regularity conditions, the statistic  $\mathbb{D}_\phi$  converge weakly to three different functions of the  $\chi_{p+q}^2$  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 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 by Uchida and Yoshida (2005).

For continuous time observations from diffusion processes, Vajda (1990) considered the model  $dX(t) = -b(t)X_t dt + \sigma(t)dW_t$ ; K uchler and S orensen (1997) and Morales *et al.* (2004) contain several results on the likelihood ratio test statistics and R enyi statistics for exponential family of diffusions. Explicit derivations of the R enyi information on the invariant law of ergodic diffusion processes have been presented in De Gregorio and Iacus (2007). 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 from general diffusion processes is still a developing stream of research. Kutoyants (2004) and Dachian and Kutoyants (2008) consider the problem of testing statistical hypotheses for ergodic diffusion models in continuous time; Kutoyants (1984) and Iacus and Kutoyants (2001) consider parametric and semiparametric hypotheses testing for small diffusion processes; Negri and Nishiyama (2007a, b) propose a non parametric test based on score marked empirical process for both continuous and discrete time observation from small diffusion processes further extended to 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. A it-Sahalia (1996, 2008), Giet and Lubrano (2008) and Chen *et al.* (2008) proposed tests based on the 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 Section 5.

## 2. Assumptions on diffusion model

We consider the family of one-dimensional diffusion processes  $\{X_t, t \in [0, T]\}$ , solution to

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

where  $W_t$  is a Brownian motion. Let  $\theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta$ , where  $\Theta_\alpha$  and  $\Theta_\beta$  are respectively compact convex subset of  $\mathbb{R}^p$  and  $\mathbb{R}^q$ . Furthermore

we assume that the drift function  $b : \mathbb{R} \times \Theta_\alpha \rightarrow \mathbb{R}$  and the diffusion coefficient  $\sigma : \mathbb{R} \times \Theta_\beta \rightarrow \mathbb{R}$  are known apart from the parameters  $\alpha$  and  $\beta$ . We assume that the process  $X_t$  is ergodic for every  $\theta$  with invariant law  $\mu_\theta$ . The process  $X_t$  is observed at discrete times  $t_i = i\Delta_n, i = 0, 1, 2, \dots, n$ , where  $\Delta_n$  is the length of the steps. We indicate the observations with  $\mathbf{X}_n = \{X_{t_i}\}_{0 \leq i \leq n}$ . The asymptotic is  $\Delta_n \rightarrow 0, n\Delta_n \rightarrow \infty$  and  $n\Delta_n^2 \rightarrow 0$  as  $n \rightarrow \infty$ .

In the definition of the  $\phi$ -divergence (1.1) the likelihood of the process is need, but as noted in the Introduction, this is usually not know. There are several ways to approximate the likelihood of a discretely observed diffusion process (for a review see, e.g., Chap. 3, Iacus, 2008). In this paper, we use the approximation proposed by Dacunha-Castelle and Florens-Zmirou (1986) although our result hold true (with some adaptations of the proofs) for other approximations, like, e.g. the one based on Hermite polynomial expansion by Ait-Sahalia (2002). To write it in explicit way, we use the same setup as in Uchida and Yoshida (2005). We introduce the following functions

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

$$\tilde{B}(x, \theta) = B(s^{-1}(\beta, x), \theta), \quad \tilde{h}(x, \theta) = \tilde{B}^2(x, \theta) + \tilde{B}'(x, \theta)$$

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 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  $\theta$  with invariant probability measure  $\mu_\theta$ . All polynomial moments of  $\mu_\theta$  are finite.*

*iv) For all  $m \geq 0$  and for all  $\theta$ ,  $\sup_t E|X_t|^m < \infty$ .*

*v) For every  $\theta$ , the coefficients  $b(\alpha, x)$  and  $\sigma(\beta, x)$  are twice differentiable with respect to  $x$  and the derivatives are polynomial growth in  $x$ , uniformly in  $\theta$ .*

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 respect to  $\theta$  for all  $x$  in the state space. All derivatives respect to  $\theta$  are polynomial growth in  $x$ , uniformly in  $\theta$ .

**Assumption 2.2.** [Regularity for the approximation]

i)  $\tilde{h}(x, \theta) = O(|x|^2)$  as  $x \rightarrow \infty$ .

ii)  $\inf_x \tilde{h}(x, \theta) > -\infty$  for all  $\theta$ .

iii)  $\sup_\theta \sup_x |\tilde{h}^3(x, \theta)| \leq M < \infty$ .

iv) There exists  $\gamma > 0$  such that for every  $\theta$  and  $j = 1, 2$ ,  $|\tilde{B}^j(x, \theta)| = O(|\tilde{B}(x, \theta)|^\gamma)$  as  $|x| \rightarrow \infty$ .

**Assumption 2.3.** [Identifiability] The coefficients  $b(\alpha, x) = b(\alpha_0, x)$  and  $\sigma(\beta, x) = \sigma(\beta_0, x)$  for  $\mu_{\theta_0}$  a.s. all  $x$  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 transition density  $f$  of the process  $X$  from  $y$  to  $x$  at lag  $t$

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

and its logarithm

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

where

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

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

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

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

$$f_n(\mathbf{X}_n, \theta) = \prod_{i=1}^n f(\Delta_n, X_{t_{i-1}}, X_{t_i}, \theta) \quad \text{and} \quad l_n(\mathbf{X}_n, \theta) = \sum_{i=1}^n l(\Delta_n, X_{t_{i-1}}, X_{t_i}, \theta)$$

### 3. Construction of the test statistics and results

Consider the divergence defined in (1.1) and let  $\phi(\cdot)$  be such that  $\phi(1) = 0$  and, when they exist, define  $C_\phi = \phi'(1)$  and  $K_\phi = \phi''(1)$ . We consider three different setup

**Assumption 3.1.**  $C_\phi \neq 0$  is a finite constant depending only on  $\phi$  and independent of  $\theta$ ;

**Assumption 3.2.**  $C_\phi = 0$  and  $K_\phi \neq 0$  is a finite constant depending only on  $\phi$  and independent of  $\theta$ ;

**Assumption 3.3.**  $C_\phi \neq 0$  and  $K_\phi \neq 0$  are finite constants depending only on  $\phi$  and independent of  $\theta$ ;

**Remark 3.1.** The above Assumptions are not so strong. In fact, for example the  $\alpha$ -divergences  $D_{\phi_\alpha}(\theta, \theta_0)$  satisfy the Assumptions 3.1 and 3.3, while for the power-divergences  $D_{\phi_\lambda}(\theta, \theta_0)$  it's easy to verify that  $C_\phi = \phi'(1) = 0$ .

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

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

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

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

We indicate with  $\Gamma$  the  $(p+q) \times (p+q)$  matrix

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

and  $I_p$  is the  $p \times p$  identity matrix. Using the approximated likelihood  $f_n(\mathbf{X}_n, \theta)$  and  $f_n(\mathbf{X}_n, \theta_0)$ , the  $\phi$ -divergence in (1.1) becomes

$$D_\phi(\theta, \theta_0) = E_{\theta_0} \phi \left( \frac{f_n(\mathbf{X}_n, \theta)}{f_n(\mathbf{X}_n, \theta_0)} \right) \quad (3.2)$$

To construct a test statistics we replace  $\theta$  by the estimator  $\tilde{\theta}_n(\mathbf{X}_n)$  and, having only one single observation of  $\mathbf{X}_n$ , i.e. only one observed trajectory, we estimate (3.2) with

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) = \phi \left( \frac{f_n(\mathbf{X}_n, \tilde{\theta}_n(\mathbf{X}_n))}{f_n(\mathbf{X}_n, \theta_0)} \right) \quad (3.3)$$

Please notice that, conversely to the i.i.d. case, there is no integral in the definition of (3.3). We will discuss this point after the presentation of the Theorem 3.1. The proposed test for testing  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta \neq \theta_0$  is realized as  $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) = 0$  versus  $\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \neq 0$ .

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

*i) if function  $\phi(\cdot)$  satisfies Assumption 3.1, then*

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \xrightarrow{d} C_\phi \chi_{p+q}^2 \quad (3.4)$$

*ii) if function  $\phi(\cdot)$  satisfies Assumption 3.2, then*

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \xrightarrow{d} \frac{K_\phi}{2} Z_{p+q} \quad (3.5)$$

where  $\sqrt{Z_{p+q}} = \chi_{p+q}^2$ .

*iii) if function  $\phi(\cdot)$  satisfies Assumption 3.3, then*

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \xrightarrow{d} \frac{1}{2} (C_\phi \chi_{p+q}^2 + (C_\phi + K_\phi) Z_{p+q}) \quad (3.6)$$

**Remark 3.2.** *It's clear that for  $C_\phi = 0$  from (3.6) we immediately reobtain the convergence result (3.5).*

**Remark 3.3.** *If we consider the limits as  $\alpha \rightarrow -1$  for  $\phi_\alpha(x)$  of the  $\alpha$ -divergences, i.e. we consider the Kullback-Leibler divergence, we have*

$$\phi(x) = \lim_{\alpha \rightarrow -1} \phi_\alpha(x) = -\log(x)$$

for which  $C_\phi = -1$  and  $K_\phi = 1$ . In that case, (3.6) reduces to the standard result for the likelihood ratio test statistics.

The convergence in Theorem 3.1 may appear somewhat strange if one thinks about the usual results on  $\phi$ -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  $D_\phi$  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 appears the expected value of the second order derivative which converges to the Fisher information and, in our case, we have not the expected value, hence the convergence to the square of the  $\chi^2$  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  $\phi$ -divergence takes the usual form of the i.i.d. case because the invariant density have the explicit form. Indeed, let

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

be the scale and speed functions of the diffusion, with  $\tilde{x}$  some value in the state space of the diffusion process. Let  $M = \int m(x, \theta) dx$ , 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  $\phi$ -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 3.4.** *In our application, to derive and estimator, we consider further the local gaussian approximation of the same transition density (see,*

Yoshida, 1992)

$$g_n(\mathbf{X}_n, \theta) = \sum_{i=1}^n g_n(\Delta_n, X_{t_{i-1}}, X_{t_i}, \theta) \quad (3.7)$$

where

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

The approximate maximum likelihood estimator  $\hat{\theta}_n(\mathbf{X}_n)$  based on (3.7) is then defined as

$$\hat{\theta}_n(\mathbf{X}_n) = \arg \sup_{\theta} g_n(\mathbf{X}_n, \theta) \quad (3.8)$$

Under the condition  $n\Delta_n^2 \rightarrow 0$  (see Theorem 1 in Kessler, 1997) the estimator  $\hat{\theta}_n(\mathbf{X}_n)$  in (3.8) satisfies (3.1). Hence, the result of Theorem 3.1 applies for  $\hat{\theta}_n(\mathbf{X}_n) = \hat{\theta}_n(\mathbf{X}_n)$ .

**Remark 3.5.** 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  $C_\phi$  and  $K_\phi$  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  $\theta$  for the drift and diffusion coefficients are different and are respectively equal to  $\sqrt{n\Delta_n}$  and  $\sqrt{n}$ .

**Remark 3.6.** As remarked in Uchida and Yoshida (2001), it is always better to derive approximate ML estimators and the test statistics on different approximations of the true likelihood to avoid circularities.

#### 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 (3.8) and the following quantities

- estimated  $\alpha$ -divergences

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

with  $\phi_\alpha(x) = 4(1 - x^{\frac{1+\alpha}{2}})/(1 - \alpha^2)$ , with  $C_\alpha = \frac{2}{\alpha-1}$  and  $K_\phi = 1$ . We consider  $\alpha \in \{-0.99, -0.90, -0.75, -0.50, -0.25, -0.10\}$ ;

- estimated 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  $\phi_\lambda(x) = (x^{\lambda+1} - x - \lambda(x - 1))/(\lambda(\lambda + 1))$ , with  $C_\lambda = 0$ ,  $K_\lambda = 1$ . We consider  $\lambda \in \{-0.99, -1.20, -1.50, -1.75, -2.00, -2.50\}$ ;

- likelihood ratio statistic

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

For  $\mathbb{D}_\alpha$  and  $\mathbb{D}_\lambda$ , the threshold of the rejection region of the test are calculated using formula (3.6) as the empirical quantiles of (3.6) of 100000 simulations of the random variable  $\chi_{p+q}^2$ . For  $\mathbb{D}_{\log}$  is again used formula (3.6) but exact quantiles of the random variable  $\chi_{p+q}^2$  are used. Because the interest is in testing  $\mathbb{D}_\phi = 0$  against  $\mathbb{D}_\phi \neq 0$ , whenever  $f_n(\mathbf{X}_n, \hat{\theta}_n(\mathbf{X}_n)) > f_n(\mathbf{X}_n, \theta_0)$  we exchange the numerator and the denominator to avoid negative signs in the test statistics. Usually, this is not going to happen if  $\phi$  is convex and  $\phi'(1) = 0$  (see, e.g. Morales *et al.*, 1997).

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  $A$ ,  $M = 10000$  is the number of simulations and  $c_\alpha$  is the  $(1 - \alpha)\%$  quantile of the proper distribution. Similarly we calculate the power of the test under alternative models as

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

In our experiments we consider the two families of stochastic processes borrowed from finance

The Vasicek (VAS) model:  $dX_t = \kappa(\alpha - X_t)dt + \sigma X_t dW_t$ , where, in finance,  $\sigma$  is interpreted as volatility,  $\alpha$  is the long-run equilibrium value of the process and  $\kappa$  is the speed of reversion. Let  $(\kappa_0, \alpha_0, \sigma_0^2) = (0.85837, 0.089102, 0.0021854)$ , we consider three different sets of hypotheses for the parameters

model	$\theta = (\kappa, \alpha, \sigma^2)$
VAS <sub>0</sub>	$(\kappa_0, \alpha_0, \sigma_0^2)$
VAS <sub>1</sub>	$(4 \cdot \kappa_0, \alpha_0, 4 \cdot \sigma_0^2)$
VAS <sub>2</sub>	$(\frac{1}{4} \kappa_0, \alpha_0, \frac{1}{4} \cdot \sigma_0^2)$

The interesting facts are that VAS<sub>0</sub>, VAS<sub>1</sub> and VAS<sub>2</sub> 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

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

and both show a strong dependency of the covariance as a function of  $\kappa$ , 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 sets of hypotheses for the parameters

model	$\theta = (\kappa, \alpha, \sigma^2)$
CIR <sub>0</sub>	$(\kappa_0, \alpha_0, \sigma_0^2)$
CIR <sub>1</sub>	$(\frac{1}{2} \cdot \kappa_0, \alpha_0, \frac{1}{2} \cdot \sigma_0^2)$
CIR <sub>2</sub>	$(\frac{1}{4} \cdot \kappa_0, \alpha_0, \frac{1}{4} \cdot \sigma_0^2)$

This model has a transition density of  $\chi^2$ -type, hence local gaussian approximation is less likely to hold for non negligible values of  $\Delta_n$ .

The parameters of the above models, have been chosen according to Pritsker (1998) and Chen *et al.* (2008), in particular VAS<sub>0</sub> corresponds to the model estimated by Ait-Sahalia (1996) for real interest rates data.

We study the level and the power of the three family of test statistics for different values of  $\Delta_n \in \{0.1, 0.001\}$  and  $n \in \{50, 100, 500\}$ . For the same trajectory, hence we simulate 1000 observations and we 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 the Tables 1-9. We point out that in the Tables 3, 5, 7 and 9, in the column “model  $(\alpha, n)$ ” the  $\alpha$  corresponds to the true level of the test used to calculate  $c_\alpha$ . The other  $\alpha$ 's in the first row of the tables correspond to the  $\alpha$  in  $\phi_\alpha$ -divergences.

*Summary of the analysis for the Vasicek model.* It turns out that  $\alpha$ -divergences are not very good in terms of estimated level of the test, but their power function behaves as expected. It also 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 4, row 2), although this is not the case for the power-divergences and the likelihood ratio test (Tables 3 and 1, row 2).

In general power divergences for  $\lambda$  in  $\{-0.99, -1.20, -1.50, -1.75, -2.00\}$  have always very small estimated level and high power under the selected alternatives. The  $\alpha$ -divergences, do not behave very good and, the way they are defined, only approximate the likelihood ratio for  $\alpha = -0.99$ .

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.* 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  $CIR_2$  doesn't present particular problems.

## 5. Proofs

The following important Lemmas are useful to prove the Theorem 3.1.

**Lemma 5.1** (Kessler, 1997). *Under the assumptions 2.1-2.3, as  $n\Delta_n^2 \rightarrow 0$  the following hold true*

$$\Gamma^{\frac{1}{2}} \nabla_{\theta} g_n(\mathbf{X}_n, \theta_0) \xrightarrow{p} N(0, \mathcal{I}(\theta_0)) \quad (5.1)$$

**Lemma 5.2** (Uchida and Yoshida, 2005). *Under the assumptions 2.1-2.3, as  $n\Delta_n^2 \rightarrow 0$  the following hold true*

$$\Gamma^{\frac{1}{2}} \nabla_{\theta} l_n(\mathbf{X}_n, \theta_0) = \Gamma^{\frac{1}{2}} \nabla_{\theta} g_n(\mathbf{X}_n, \theta_0) + o_p(1) \quad (5.2)$$

**Lemma 5.3** (Uchida and Yoshida, 2005). *Under the assumptions 2.1-2.3, as  $n\Delta_n^2 \rightarrow 0$  the following hold true*

$$\Gamma^{\frac{1}{2}} \nabla_{\theta}^2 l_n(\mathbf{X}_n, \theta_0) \Gamma^{\frac{1}{2}} \xrightarrow{p} -\mathcal{I}(\theta_0) \quad (5.3)$$

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

i) We can write that

$$\begin{aligned} \mathbb{D}_{\phi}(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) &= \mathbb{D}_{\phi}(\theta_0, \theta_0) + [\nabla_{\theta} \mathbb{D}_{\phi}(\theta_0, \theta_0)]^T (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) + o_p(1) \\ &= [\nabla_{\theta} \mathbb{D}_{\phi}(\theta_0, \theta_0)]^T (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) + o_p(1) \end{aligned}$$

because  $\mathbb{D}_{\phi}(\theta_0, \theta_0) = 0$ . Noting that for  $k = 1, \dots, p + q$

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

by Assumption 3.1 follows that

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

and therefore

$$\mathbb{D}_{\phi}(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) = C_{\phi} \left[ \Gamma^{\frac{1}{2}} \nabla_{\theta} l_n(\mathbf{X}_n, \theta_0) \right]^T \Gamma^{-\frac{1}{2}} (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) + o_p(1) \quad (5.4)$$

From (5.4) by means of Lemma 5.2-5.1 and Slutsky's Theorem immediately follows

$$\mathbb{D}_{\phi}(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \xrightarrow{d} C_{\phi} \chi_{p+q}^2$$

ii) Since for  $k, j = 1, \dots, p + q$

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

follows that

$$\begin{aligned} \mathbb{D}_{\phi}(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) &= \frac{1}{2} [\Gamma^{-1/2} (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0)]^T \Gamma^{1/2} \nabla_{\theta}^2 \mathbb{D}_{\phi}(\theta_0, \theta_0) \Gamma^{1/2} \\ &\quad \times \Gamma^{-1/2} (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) + o_p(1) \\ &= \frac{K_{\phi}}{2} [\Gamma^{-1/2} (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0)]^T \Gamma^{1/2} \nabla_{\theta} l_n(\mathbf{X}_n, \theta_0) [\Gamma^{1/2} \nabla_{\theta} l_n(\mathbf{X}_n, \theta_0)]^T \\ &\quad \times \Gamma^{-1/2} (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) + o_p(1) \end{aligned}$$

From (5.4) by means of Lemma 5.2-5.1 and Slutsky's Theorem immediately follows

$$\mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) \xrightarrow{d} \frac{K_\phi}{2} Z_{p+q}$$

It's easy to verify that the density function of the r.v.  $Z_{p+q}$  is equal to

$$f_{Z_{p+q}}(z) = \frac{(1/2)^{\frac{p+q}{2}}}{\Gamma\left(\frac{p+q}{2}\right)} \sqrt{z}^{\frac{p+q}{2}-1} e^{-\sqrt{z}/2} \frac{1}{2\sqrt{z}}, \quad z > 0 \quad (5.5)$$

iii) By previous considerations we have that

$$\begin{aligned} \mathbb{D}_\phi(\tilde{\theta}_n(\mathbf{X}_n), \theta_0) &= [\nabla_\theta \mathbb{D}_\phi(\theta_0, \theta_0)]^T (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) \\ &\quad + \frac{1}{2} [\Gamma^{-1/2} (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0)]^T \Gamma^{1/2} \nabla_\theta^2 \mathbb{D}_\phi(\theta_0, \theta_0) \Gamma^{1/2} \\ &\quad \times \Gamma^{-1/2} (\tilde{\theta}_n(\mathbf{X}_n) - \theta_0) + o_p(1) \end{aligned} \quad (5.6)$$

where

$$\begin{aligned} &\nabla_\theta^2 \mathbb{D}_\phi(\theta_0, \theta_0) \\ &= K_\phi \nabla_\theta l_n(\mathbf{X}_n, \theta_0) [\nabla_\theta l_n(\mathbf{X}_n, \theta_0)]^T + C_\phi \frac{1}{f(\mathbf{X}_n, \theta_0)} \nabla_\theta^2 f(\mathbf{X}_n, \theta_0) \\ &= (K_\phi + C_\phi) \nabla_\theta l_n(\mathbf{X}_n, \theta_0) [\nabla_\theta l_n(\mathbf{X}_n, \theta_0)]^T + C_\phi \nabla_\theta^2 l_n(\mathbf{X}_n, \theta_0) \end{aligned} \quad (5.7)$$

Plugging in (5.6) the quantity (5.7) we derive, applying again Lemma 5.1-5.3, the following result

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

□

## Conclusions

It seems that, as in the i.i.d. case, also for discretely observed diffusion processes the  $\phi$ -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, 2008) and freely available at <http://cran.R-Project.org> contains the function `sdeDiv` which implements the  $\phi$ -divergence test statistics.

## References

- Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate, *Rev. Financial 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.
- Aït-Sahalia, Y., Fan, J., Peng, H. (2006) Nonparametric transition based tests for jump diffusions. Working paper. Available at <http://www.princeton.edu/yacine/research.htm>
- Amari, S.(1985) *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statist., Vol. **28**, New York, Springer.
- Beran, R.J. (1977) Minimum Hellinger estimates for parametric models, *Annals of Statistics*, **5**, 445-463.
- Burbea, J. (1984) The convexity with respect to gaussian distribution of divergences of order  $\alpha$ , *Utilitas Math.*, **26**, 171-192.
- Chandra, S.A., Taniguchi, M. (2006) Minimum  $\alpha$ -divergence estimation for arch models, *Journal of Time Series Analysis*, **27**(1), 19-39.
- Chen, S.X., Gao, J., Cheng, Y.T. (2008) A test for model specification of diffusion processes, *Ann. Stat.*, **36**(1), 167-198.
- Cressie, N., Read, T.R.C. (1984) Multinomial goodness of fit tests, *J. Roy. Statist. Soc. Ser B*, **46**, 440-464.
- Csiszár, I. (1967) On topological properties of  $f$ -divergences, *Studia Scientifica Mathematicae Hungarica*, **2**, 329-339.
- Dachian, S., Kutoyants, Yu. A. (2008) On the goodness-of-t tests for some continuous time processes, in *Statistical Models and Methods for Biomedical and Technical Systems*, 395-413. Vonta, F., Nikulin, M., Limnios, N. and Huber-Carol C. (Eds), Birkhuser, Boston.
- 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. (2007) Rényi information for ergodic diffusion processes. Available at <http://arxiv.org/abs/0711.1789>
- 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, *Computational Statistics and Data Analysis*, **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, *Mathematical Methods of Statistics*, **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-Verlag, 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, *Journal of Korean Mathematical Society*, **45**(3), 2008.
- Liese, F., Vajda, I. (1987) *Convex Statistical Distances*, Tuebner, Leipzig.
- Morales, D., Pardo, L., Vajda, I. (1997) Some New Statistics for Testing Hypotheses in Parametric Models, *Journal of Multivariate Analysis*, **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. (2007a) Goodness of fit test for ergodic diffusion processes, to appear in *Ann. Inst. Statist. Math.*. Available at <http://arxiv.org/abs/math/0701022v1>

- Negri, I. and Nishiyama, Y. (2007b). Goodness of t test for small diffusions based on discrete observations, *Research Memorandum 1054, Inst. Statist. Math.*, Tokyo.
- 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.
- Pritsker, M. (1998) Nonparametric density estimation and tests of continuous time interest rate models, *Rev. financial Studies*, **11**, 449-487.
- R Development Core Team (2008) *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, *Proceedings of the Fourth Berkeley Symposium on Probability and Mathematical Statistics*, Vol 1, University of California, Berkeley, 547-461.
- Rivas, M.J., Santos, M.T., Morales, D. (2005) Rényi test statistics for partially observed diffusion processes, *Journal of Statistical Planning and 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, *Statistical Inference for Stochastic Processes*, **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.
- Yoshida, N. (1992) Estimation for diffusion processes from discrete observation, *J. Multivar. Anal.*, **41**(2), 220-242.

model (n)	$\alpha = 0.01$	$\alpha = 0.05$	model (n)	$\alpha = 0.01$	$\alpha = 0.05$
VAS <sub>0</sub> (50)	0.01	0.04	VAS <sub>0</sub> (50)	0.01	0.04
VAS <sub>1</sub> (50)	1.00	1.00	VAS <sub>1</sub> (50)	1.00	1.00
VAS <sub>2</sub> (50)	1.00	1.00	VAS <sub>2</sub> (50)	1.00	1.00
VAS <sub>0</sub> (100)	0.01	0.04	VAS <sub>0</sub> (100)	0.01	0.04
VAS <sub>1</sub> (100)	1.00	1.00	VAS <sub>1</sub> (100)	1.00	1.00
VAS <sub>2</sub> (100)	1.00	1.00	VAS <sub>2</sub> (100)	1.00	1.00
VAS <sub>0</sub> (500)	0.01	0.07	VAS <sub>0</sub> (500)	0.00	0.02
VAS <sub>1</sub> (500)	1.00	1.00	VAS <sub>1</sub> (500)	1.00	1.00
VAS <sub>2</sub> (500)	1.00	1.00	VAS <sub>2</sub> (500)	1.00	1.00

Table 1: Numbers represent probability of rejection under the true generating model, with  $c_\alpha$  calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Likelihood ratio, for  $\Delta_n = 0.001$  (left) and  $\Delta_n = 0.1$  (right).

model (n)	$\alpha = 0.01$	$\alpha = 0.05$	model (n)	$\alpha = 0.01$	$\alpha = 0.05$
CIR <sub>0</sub> (50)	0.02	0.11	CIR <sub>0</sub> (50)	0.01	0.04
CIR <sub>1</sub> (50)	0.59	0.84	CIR <sub>1</sub> (50)	0.78	0.93
CIR <sub>2</sub> (50)	1.00	1.00	CIR <sub>2</sub> (50)	1.00	1.00
CIR <sub>0</sub> (100)	0.03	0.11	CIR <sub>0</sub> (100)	0.01	0.04
CIR <sub>1</sub> (100)	0.96	0.99	CIR <sub>1</sub> (100)	0.99	1.00
CIR <sub>2</sub> (100)	1.00	1.00	CIR <sub>2</sub> (100)	1.00	1.00
CIR <sub>0</sub> (500)	0.02	0.09	CIR <sub>0</sub> (500)	0.00	0.02
CIR <sub>1</sub> (500)	1.00	1.00	CIR <sub>1</sub> (500)	1.00	1.00
CIR <sub>2</sub> (500)	1.00	1.00	CIR <sub>2</sub> (500)	1.00	1.00

Table 2: Numbers represent probability of rejection under the true model, with rejection region calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Likelihood ratio, for  $\Delta_n = 0.001$  (left) and  $\Delta_n = 0.1$  (right).

model ( $\alpha, n$ )	$\alpha = -0.99$	$\alpha = -0.90$	$\alpha = -0.75$	$\alpha = -0.50$	$\alpha = -0.25$	$\alpha = -0.10$
VAS <sub>0</sub> (0.01, 50)	0.01	0.10	0.39	0.62	0.73	0.77
VAS <sub>1</sub> (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 50)	0.04	0.12	0.39	0.62	0.73	0.77
VAS <sub>1</sub> (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.01, 100)	0.01	0.10	0.39	0.63	0.74	0.78
VAS <sub>1</sub> (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 100)	0.04	0.11	0.40	0.63	0.74	0.78
VAS <sub>1</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.01, 500)	0.02	0.18	0.61	0.83	0.90	0.92
VAS <sub>1</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 500)	0.07	0.20	0.61	0.83	0.90	0.92
VAS <sub>1</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00

Table 3: Numbers represent probability of rejection under the true generating model, with  $c_\alpha$  calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Estimated  $\alpha$ -divergences, for  $\Delta_n = 0.001$ .

model ( $\alpha, n$ )	$\lambda = -0.99$	$\lambda = -1.20$	$\lambda = -1.50$	$\lambda = -1.75$	$\lambda = -2.00$	$\lambda = -2.50$
VAS <sub>0</sub> (0.01, 50)	0.00	0.00	0.00	0.01	0.02	0.04
VAS <sub>1</sub> (0.01, 50)	0.00	0.99	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 50)	0.40	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 50)	0.00	0.00	0.00	0.01	0.03	0.06
VAS <sub>1</sub> (0.05, 50)	0.67	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 50)	0.99	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.01, 100)	0.00	0.00	0.00	0.01	0.02	0.04
VAS <sub>1</sub> (0.01, 100)	0.23	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 100)	0.88	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 100)	0.00	0.00	0.00	0.01	0.03	0.06
VAS <sub>1</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.01, 500)	0.00	0.00	0.00	0.01	0.03	0.08
VAS <sub>1</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 500)	0.00	0.00	0.01	0.03	0.06	0.12
VAS <sub>1</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00

Table 4: Numbers represent probability of rejection under the true generating model, with  $c_\alpha$  calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Estimated power-divergences for  $\Delta_n = 0.001$

model ( $\alpha, n$ )	$\alpha = -0.99$	$\alpha = -0.90$	$\alpha = -0.75$	$\alpha = -0.50$	$\alpha = -0.25$	$\alpha = -0.10$
VAS <sub>0</sub> (0.01, 50)	0.01	0.15	0.55	0.78	0.86	0.88
VAS <sub>1</sub> (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 50)	0.05	0.17	0.55	0.78	0.86	0.88
VAS <sub>1</sub> (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.01, 100)	0.01	0.13	0.48	0.71	0.80	0.83
VAS <sub>1</sub> (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 100)	0.04	0.15	0.48	0.71	0.80	0.83
VAS <sub>1</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.01, 500)	0.00	0.06	0.25	0.54	0.69	0.74
VAS <sub>1</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 500)	0.02	0.07	0.25	0.54	0.69	0.74
VAS <sub>1</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00

Table 5: Numbers represent probability of rejection under the true generating model, with  $c_\alpha$  calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Estimated  $\alpha$ -divergences, for  $\Delta_n = 0.1$

model ( $\alpha, n$ )	$\lambda = -0.99$	$\lambda = -1.20$	$\lambda = -1.50$	$\lambda = -1.75$	$\lambda = -2.00$	$\lambda = -2.50$
VAS <sub>0</sub> (0.01, 50)	0.00	0.00	0.00	0.01	0.02	0.05
VAS <sub>1</sub> (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 50)	0.00	0.00	0.00	0.02	0.03	0.09
VAS <sub>1</sub> (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.01, 100)	0.00	0.00	0.00	0.00	0.01	0.05
VAS <sub>1</sub> (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 100)	0.00	0.00	0.00	0.01	0.03	0.08
VAS <sub>1</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.01, 500)	0.00	0.00	0.00	0.00	0.01	0.02
VAS <sub>1</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>0</sub> (0.05, 500)	0.00	0.00	0.00	0.01	0.01	0.04
VAS <sub>1</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00
VAS <sub>2</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00

Table 6: Numbers represent probability of rejection under the true generating model, with  $c_\alpha$  calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Estimated power-divergences for  $\Delta_n = 0.1$

model ( $\alpha, n$ )	$\alpha = -0.99$	$\alpha = -0.90$	$\alpha = -0.75$	$\alpha = -0.50$	$\alpha = -0.25$	$\alpha = -0.10$
CIR <sub>0</sub> (0.01, 50)	0.03	0.28	0.69	0.83	0.89	0.90
CIR <sub>1</sub> (0.01, 50)	0.63	0.95	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 50)	0.12	0.31	0.69	0.83	0.89	0.90
CIR <sub>1</sub> (0.05, 50)	0.86	0.96	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.01, 100)	0.03	0.28	0.69	0.85	0.89	0.91
CIR <sub>1</sub> (0.01, 100)	0.97	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 100)	0.12	0.31	0.69	0.85	0.89	0.91
CIR <sub>1</sub> (0.05, 100)	0.99	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.01, 500)	0.03	0.22	0.59	0.79	0.86	0.88
CIR <sub>1</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 500)	0.09	0.24	0.59	0.79	0.86	0.89
CIR <sub>1</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00

Table 7: Numbers represent probability of rejection under the true model, with rejection region calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Estimated  $\alpha$ -divergences, for  $\Delta_n = 0.001$

model ( $\alpha, n$ )	$\lambda = -0.99$	$\lambda = -1.20$	$\lambda = -1.50$	$\lambda = -1.75$	$\lambda = -2.00$	$\lambda = -2.50$
CIR <sub>0</sub> (0.01, 50)	0.00	0.00	0.00	0.02	0.05	0.13
CIR <sub>1</sub> (0.01, 50)	0.00	0.01	0.28	0.55	0.71	0.87
CIR <sub>2</sub> (0.01, 50)	0.00	0.81	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 50)	0.00	0.00	0.01	0.04	0.09	0.19
CIR <sub>1</sub> (0.05, 50)	0.00	0.09	0.48	0.70	0.81	0.92
CIR <sub>2</sub> (0.05, 50)	0.00	0.97	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.01, 100)	0.00	0.00	0.00	0.02	0.05	0.14
CIR <sub>1</sub> (0.01, 100)	0.00	0.21	0.83	0.95	0.98	0.99
CIR <sub>2</sub> (0.01, 100)	0.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 100)	0.00	0.00	0.01	0.05	0.09	0.20
CIR <sub>1</sub> (0.05, 100)	0.00	0.53	0.93	0.98	0.99	1.00
CIR <sub>2</sub> (0.05, 100)	0.24	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.01, 500)	0.00	0.00	0.00	0.02	0.04	0.10
CIR <sub>1</sub> (0.01, 500)	0.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 500)	0.00	0.00	0.01	0.04	0.07	0.15
CIR <sub>1</sub> (0.05, 500)	0.95	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00

Table 8: Numbers represent probability of rejection under the true model, with rejection region calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Estimated power-divergences for  $\Delta_n = 0.001$

model ( $\alpha, n$ )	$\alpha = -0.99$	$\alpha = -0.90$	$\alpha = -0.75$	$\alpha = -0.50$	$\alpha = -0.25$	$\alpha = -0.10$
CIR <sub>0</sub> (0.01, 50)	0.01	0.14	0.54	0.77	0.85	0.87
CIR <sub>1</sub> (0.01, 50)	0.80	0.98	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.01, 50)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 50)	0.05	0.16	0.54	0.77	0.85	0.87
CIR <sub>1</sub> (0.05, 50)	0.94	0.98	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.05, 50)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.01, 100)	0.01	0.13	0.49	0.71	0.79	0.82
CIR <sub>1</sub> (0.01, 100)	0.99	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.01, 100)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 100)	0.04	0.15	0.49	0.71	0.79	0.82
CIR <sub>1</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.05, 100)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.01, 500)	0.00	0.06	0.28	0.54	0.69	0.74
CIR <sub>1</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 500)	0.02	0.08	0.28	0.54	0.69	0.74
CIR <sub>1</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00

Table 9: Numbers represent probability of rejection under the true model, with rejection region calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Estimated  $\alpha$ -divergences, for  $\Delta_n = 0.1$

model ( $\alpha, n$ )	$\lambda = -0.99$	$\lambda = -1.20$	$\lambda = -1.50$	$\lambda = -1.75$	$\lambda = -2.00$	$\lambda = -2.50$
CIR <sub>0</sub> (0.01, 50)	0.00	0.00	0.00	0.01	0.02	0.06
CIR <sub>1</sub> (0.01, 50)	0.00	0.06	0.52	0.75	0.86	0.94
CIR <sub>2</sub> (0.01, 50)	0.00	0.99	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 50)	0.00	0.00	0.00	0.02	0.04	0.09
CIR <sub>1</sub> (0.05, 50)	0.00	0.23	0.70	0.85	0.92	0.96
CIR <sub>2</sub> (0.05, 50)	0.06	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.01, 100)	0.00	0.00	0.00	0.01	0.02	0.05
CIR <sub>1</sub> (0.01, 100)	0.00	0.56	0.96	0.99	1.00	1.00
CIR <sub>2</sub> (0.01, 100)	0.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 100)	0.00	0.00	0.00	0.02	0.03	0.08
CIR <sub>1</sub> (0.05, 100)	0.00	0.83	0.99	1.00	1.00	1.00
CIR <sub>2</sub> (0.05, 100)	0.97	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.01, 500)	0.00	0.00	0.00	0.00	0.01	0.02
CIR <sub>1</sub> (0.01, 500)	0.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.01, 500)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>0</sub> (0.05, 500)	0.00	0.00	0.00	0.01	0.02	0.04
CIR <sub>1</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00
CIR <sub>2</sub> (0.05, 500)	1.00	1.00	1.00	1.00	1.00	1.00

Table 10: Numbers represent probability of rejection under the true model, with rejection region calculated under  $H_0$ . Therefore, the values are  $\hat{\alpha}$  under model “0” and  $\hat{\beta}$  otherwise. Estimates calculated on 10000 experiments. Estimated power-divergences for  $\Delta_n = 0.1$