UNIVERSITÀ DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
E METODI MATEMATICI°

QUADERNI DEL DIPARTIMENTO

Department of Information Technology and Mathematical Methods

Working Paper

Series "*Mathematics and Statistics*"

n. 11/MS – 2009

Review on goodness of fit tests for ergodic diffusion processes by different sampling schemes

by

Ilia Negri, Yoichi Nishiyama

° Viale Marconi. 5, I – 24044 Dalmine (BG), ITALY, Tel. +39-035-2052339; Fax. +39-035-562779

Review on goodness of fit tests for ergodic diffusion processes by different sampling schemes *

Ilia Negri[†] Yoichi Nishiyama[‡]

Abstract

We consider a nonparametric goodness of fit test problem for the drift coefficient of one-dimensional ergodic diffusions, where the diffusion coefficient is a nuisance function which is estimated in some sense. Using a theory for the continuous observation case, we first consider a test based on discrete time observations of the processes.Then we also construct a test based on the data observed discretely in space, that is, the so-called tick time sampled data. We prove that in both sampling schemes the limit distribution of the test is the supremum of the standard Brownian motion, thus the test is asymptotically distribution free. We also show that the tests are consistent under any fixed alternatives.

Key words: consistent test, discrete time sample, tick time sample.
2000 MSC: 60G10; 60G35; 62M02.

*This work has been partially supported by MIUR 2006 Grant.

[†]Corresponding author: University of Bergamo, Viale Marconi, 5, 24044, Dalmine (BG), Italy. ilia.negri@unibg.it.

[‡]The Institute of Statistical Mathematics, 4-6-7, Minami-Azabu, Minato-ku, Tokyo, 106-8569, Japan. nisiyama@ism.ac.jp

1 Introduction

In this paper we review some recent results on goodness of fit test for ergodic diffusion process. These models are widely used in finance to model e.g. asset prices, interest or exchange rates. Despite the fact that in the last thirty years diffusion models have been proved to be immensely useful, not only in finance and more generally in economics science, but also in other fields such as biology, medicine, physics and engineering, the problem of goodness of fit tests for diffusion processes has still been a new issue in recent years. Goodness of fit tests play an important role in theoretical and applied statistics, and the study for them has a long history. Such tests are really useful especially if they are *distribution free*, in the sense that their distributions do not depend on the underlying model. The origin goes back to the Kolmogorov-Smirnov and Crámer-von Mises tests in the i.i.d. case, established early in the 20th century, which are *asymptotically distribution free*.

Although their great importance in application, the theory of goodness of fit tests for diffusion processes has not received much attention from researcher as the theory of estimation has. When the diffusion processes is observed on *continuous time* Kutoyants [17] discusses some possibilities of the construction of such tests in his Section 5.4, where he considers the Kolmogorov-Smirnov statistics based on the continuous observation of a diffusion process. The goodness of fit test based on the Kolmogorov-Smirnov statistics is asymptotically consistent and the asymptotic distribution under the null hypothesis follows from the weak convergence of the empirical process to a suitable Gaussian process. Unfortunately these tests are not asymptotically distribution free. Note that the Kolmogorov-Smirnov statistics for ergodic diffusion process was studied in Furnie [8], see also Furnie and Kutoyants [9] for more details, while the weak convergence of the empirical process was proved in Negri [20], (see Van der Vaart and Van Zanten, [28] for further developments). More recently Dachian and Kutoyants in [3] and [4] propose a modification of the Kolmogorov-Smirnov and the Cramer-von Mises test and prove that they are asimptotically distribution free and discuss the behaviour of the power function under local alternatives. Negri and Nishiyama [21] following another approach, based on the *score marked empirical processes* propose an asymptotic distribution free and consistent test. In Lee and Wee [14] a similar approach based on the residual empirical process was proposed to study a goodness of fit test for diffusion process with the drift in a parametric form and a constant diffusion coefficient. The work of Negri and Nishiyama [21] follow an approach based on the innovation martingale and it is motivated by the work of Koul and Stute [16] who considered a non-linear parametric time series model (see also Section 7.3 of Nishiyama [25] which is a reprint of his thesis in 1998). They studied the large sample behavior of the proposed test statistics under the null hypotheses and present a martingale transformation of the underlying process that makes tests based on it asymptotically distribution free. Some considerations on consistency have also been done. The approach is well expounded in Koul [15]. See Delgado and Stute [5] and references therein for more recent information.

Of course more interesting for applications are tests based on discrete time observations. In a pioneer work At-Sahalia [1] considered the problem of testing the parametric specification of an ergodic diffusion process via the transitional density of the process. Chen and Gao [2] (see also reference there in) propose a test for model specication of a parametric diffusion process based on a kernel estimation of the transitional density of the process.

In a non-parametric framework the work of Negri and Nishiyama [21] can be extended in the case with discrete time observation (Masuda and *al.* [19]) and in the case of discrete observation in space (the so called *tick time sampling*, see Negri and Nishiyama [23]). In Nishiyama [26] the same approach is used to propose two type of goodness of fit test based on a kernel estimator. The first type is for the drift and the second one is for the diffusion coefficient.

In this work we review in a unique context the tests based on the score marked empirical process to see how they works in different sample schemes. The test is for the drift coefficient while the unknown diffusion coefficient is considered as a nuisance function which is estimated in the test procedure.

In the next section we present the model, its properties and some general conditions and assumptions used through all the text. In Section 3 we present the case of continuous observation. Section 4 is devoted to the study of the test statistics when the process is observed in discrete time. In Section 5 we study the test statistics when the process is observed discretely in space, the tick-time sample scheme. Finally Section 6 is devoted to some conclusions and comments

2 Model and general conditions

Given a general stochastic basis, that is, a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and a filtration $\{\mathcal{A}_t\}_{t\geq 0}$ of $\mathcal{A}$, let us consider a one dimensional diffusion process X solution of a stochastic differential equation, that is a strong Markov process with continuous sample paths which satisfies the following

$$X_t = X_0 + \int_0^t S(X_t)dt + \int_0^t \sigma(X_t)dW_t, \tag{1}$$

where S and σ are functions which satisfy some properties described later and in the next sections, $\{W_t : t \geq 0\}$ is a standard Wiener process and the initial value X_0 is finite almost surely and is independent of W_t, $t \geq 0$.

We consider a case where a unique strong solution X of (1) exists, and we shall assume that X is ergodic. We are interested in goodness of fit test for the drift coefficient S, while the diffusion coefficient σ^2 is an unknown nuisance function which we estimate in our testing procedure. That is, we consider the problem of testing the null hypothesis $H_0 : S = S_0$ versus $H_1 : S \neq S_0$ for a given S_0. The meaning of the alternatives "$S \neq S_0$" will be precisely stated later.

The *scale function* of a diffusion process solution of the stochastic differential equation (1) is defined by

$$p(x) = \int_0^x \exp\left\{-2\int_0^y \frac{S(v)}{\sigma^2(v)}\,\mathrm{d}v\right\}\mathrm{d}y.$$

The *speed measure* of the diffusion process (1) is defined by $m_{S,\sigma}(\mathrm{d}x) = \frac{1}{\sigma(x)^2 p'(x)}\mathrm{d}x$.

By the ergodic property (see for example Gikhman and Skorohod, [11] or Durrett, [6]) there exists an unique invariant probability measure $\mu_{S,\sigma}$ such that for every measurable function $g \in L_1(\mu_{S,\sigma})$ we have with probability one,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T g(X_t)\mathrm{d}t = \int_{\mathbb{R}} g(z)\mu_{S,\sigma}(\mathrm{d}z).$$

Moreover the invariant measure $\mu_{S,\sigma}$ has a density given by

$$f_{S,\sigma}(y) = \frac{1}{m_{S,\sigma}(\mathbb{R})\sigma(y)^2}\exp\left\{2\int_0^y \frac{S(v)}{\sigma(v)^2}\mathrm{d}v\right\},$$

where

$$m_{S,\sigma}(\mathbb{R}) = \int_{-\infty}^{\infty} \frac{1}{\sigma(x)^2} \exp\left\{ 2 \int_0^x \frac{S(v)}{\sigma^2(v)} \, \mathrm{d}v \right\} \, \mathrm{d}x$$

is finite.

Let us introduce the following notation. Let $(\mathbb{T}, \rho)$ be a metric space. We denote by $C_\rho(\mathbb{T})$ the space of continuous functions on $\mathbb{T}$, by $\ell^\infty_\rho(\mathbb{T})$ the space of bounded functions on $\mathbb{T}$. We equip both the spaces with the uniform metric and we denote by "$\rightarrow^p$" and "$\rightarrow^d$" the convergence in probability and in distribution respectively. In our work we consider $\mathbb{T} = [-\infty, +\infty]$.

3 Continuous observation

Suppose that we observe the process solution of the stochastic differential equation (1) up to time T and we wish to test the two different simple hypotheses

$$H_0 : S = S_0$$
$$H_1 : S \neq S_0,$$

where $S \neq S_0$ means

$$\int_{-\infty}^{x_S} (S(y) - S_0(y)) f_{S,\sigma}(y) \mathrm{d}y \neq 0, \quad \text{for some } x \in \mathbb{R}$$

Let us introduce the following conditions

A1. The diffusion process X, which is a solution of (1) for (S, σ), is regular, and the speed measure $m_{S,\sigma}$ is finite. (Thus the process X is ergodic.)

A2. Define $\Sigma^2_{S,\sigma} = \int\limits_{-\infty}^{+\infty} \sigma^2(z) f_{S,\sigma}(z) dz$. The invariant density $f_{S,\sigma}$ satisfies $0 < \Sigma^2_{S,\sigma} < +\infty$.

We consider the stochastic process defined, for every $x \in [-\infty, +\infty]$, as

$$V_T(x) = \frac{1}{\sqrt{T}} \int_0^T \mathbf{1}_{(-\infty,x]}(X_t)(\mathrm{d}X_t - S_0(X_t)\mathrm{d}t). \tag{2}$$

Here $\mathbf{1}_A$ is the indicator function of a measurable set A. Negri and Nishiyama [21], following Koul and Stute [16], called a slightly different version of this process the *score marked empirical process*, and obtained the following result, which is a fruit of the combination of the weak convergence theory for for ℓ^∞-valued continuous martingales based on the metric entropy developed by Nishiyama [24], [25] and a theorem for local time of ergodic diffusion processes given by van Zanten [29] (see also van der Vaart and van Zanten [28]).

Theorem 1. *Assume* **A1** *and* **A2** *for* (S_0, σ)*. Under* $H_0 : S = S_0$*, suppose that a positive consistent estimator* $\widehat{\Sigma}^T$ *for* $\Sigma_{S_0,\sigma}$ *is given. Then it holds that*

$$\mathcal{C}^T = \frac{\sup_{x\in[-\infty,\infty]} |V^T(x)|}{\widehat{\Sigma}^T} \to^d \sup_{t\in[0,1]} |B_t|,$$

where $\{B_t : t \geq 0\}$ *is a standard Brownian motion.*

The proof can be found in Negri and Nishiyama [21]. As a consequence the test based on the following statistical decision function $\phi_T^* = \mathbf{1}_{\{\mathcal{C}^T > c_\alpha\}}$, where the *critical value* c_α is defined by $\mathbf{P}\left(\sup_{0\leq t\leq 1} |B_t| > c_\alpha\right) = \alpha$, is asymptotically distribution free. We observe that in continuous time observation we can always suppose that σ is known because it can be estimated without error. See Kutoyants [17]. It is well known that the distribution function of the supremum over $t \in [0, 1]$ of $|B_t|$ is given by

$$F(x) = \mathbf{P}\left(\sup_{t\in[0,1]} |B_t| \leq x\right) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \exp\left(-\frac{(2n+1)^2\pi^2}{8x^2}\right)$$

See Feller [7].

The proposed test is consistent under any fixed alternative. In order to have consistency we have to study the asymptotic behavior of $\mathcal{C}^T$ under a fixed alternative. Denote by $\mathcal{S}$ the class of function S which satisfy **A1**, **A2** and

$$\int_{-\infty}^{x_S} (S(z) - S_0(z)) f_{S,\sigma}(z)\, dz \neq 0 \quad \text{for some} \quad x_S \in (-\infty, +\infty]$$

We have the following result

Theorem 2. *Assume* **A1** *and* **A2** *for* (S, σ). *Under* $H_1 : S \in \mathcal{S}$, *if* $\widehat{\Sigma}^T$ *is bounded in probability, then*

$$\mathcal{C}^T = \frac{\sup_{x\in[-\infty,\infty]} |V^T(x)|}{\widehat{\Sigma}^T} \neq O_P(1).$$

Also the proof of this theorem can be found in Negri and Nishiyama [21].

4 Discrete time observations

The test procedure presented in the previous Section 3 is based on the continuous observation of the process on $[0, T]$. One interesting points was to try to extend the result obtained for continuous time observation when the test statistics is based on *discrete time observation*, which is more realistic in applications.

As well as Negri and Nishiyama [22], where the small diffusion model was considered, we take an approach based on a certain marked empirical process to construct an asymptotically distribution free test, where "empirical process" actually means an innovation martingale. What follows is an attempt to develope the method used in continuous time observation when we have discrete time observations.

In the sequel we denote Log $m = \log(1+m)$ and we consider the following.

Sampling scheme. The process $X = \{X_t : t \in [0, \infty)\}$ is observed at times $0 = t_0^n < t_1^n < \cdots < t_n^n$ such that, as $n \to \infty$, $t_n^n \to \infty$ and $h_n = O(n^{-2/3}(\text{Log } n)^{1/3})$ (which implies $nh_n^2 \to 0$) where $h_n = \max_{1 \le i \le n} |t_i^n - t_{i-1}^n|$.

This condition implies $h_n \to 0$, so we may assume that $h_n \le 1$ without loss of generality. We will propose an asymptotically distribution free test based on this sampling scheme, namely, *high frequency data*. We should mention that there is a huge literature on discrete time approximations of statistical estimators for diffusion processes; see e.g. the Introduction of Gobet *et al.* [12] for a review including not only high frequency cases but also low frequency cases. However, it seems difficult to obtain asymptotically distribution free results based on low frequency data.

To define the test statistics, we introduce an array of constants in the state space of the process, $-\infty = x_0^n < x_1^n < x_2^n < \cdots < x_{m(n)}^n < x_{m(n)+1}^n = \infty$ such that, as $n \to \infty$,

$$\max_{2 \le k \le m(n)} |x_k^n - x_{k-1}^n| \to 0, \quad x_1^n \downarrow -\infty, \quad x_{m(n)}^n \uparrow \infty.$$

For example, one may consider $x_k^n = -n + (k/n)$ with $k = 1, 2, ..., 2n^2$. Next we introduce a sequence of functions $\{\psi_k^n\}_{k,n}$ defined on $(-\infty, \infty)$ which approximates the indicator function $1_{(-\infty, x_k^n]}$.

Definition 1. *Let a sequence of positive constants b_n be given. For every $k = 1, 2, ..., m(n)$, ψ_k^n is the continuous, piecewise linear function on $(-\infty, \infty)$ defined by*

$$\psi_k^n(z) = \begin{cases} 1, & z \in (-\infty, x_k^n], \\ \textit{line}, & z \in [x_k^n, x_k^n + b_n], \\ 0, & z \in [x_k^n + b_n, \infty). \end{cases}$$

Also we define $\psi_0^n \equiv 0$ *and* $\psi_{m(n)+1}^n \equiv 1$.

This function satisfies the following properties:

$$|\psi_k^n(z) - \psi_k^n(z')| \leq b_n^{-1}|z - z'|;$$

$$|\psi_k^n(z) - 1_{(-\infty, x_k^n]}(z)| \leq 1_{[x_k^n, x_k^n + b_n]}(z).$$

Our test statistics is based on the random field $U^n = \{U^n(x) : x \in [-\infty, \infty]\}$ defined by $U^n(-\infty) = 0$ and

$$U^n(x) = \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^{n} \psi_k^n(X_{t_{i-1}^n})[X_{t_i^n} - X_{t_{i-1}^n} - S_0(X_{t_{i-1}^n})|t_i^n - t_{i-1}^n|]$$

for $x \in (x_{k-1}^n, x_k^n],\ 1 \leq k \leq m(n) + 1$.

We call it the *smoothed score marked empirical process* based on discrete time observation.

This U^n is an approximation of the random field $V^n = \{V^n(x) : x \in [-\infty, \infty]\}$ defined by

$$V^n(x) = \frac{1}{\sqrt{t_n^n}} \int_0^{t_n^n} 1_{(-\infty, x]}(X_t)[dX_t - S_0(X_t)dt],$$

which is the *score marked empirical process* based on continuous time observation given by (2).

Let us introduce some more conditions. First of all we observe that in order to prove the results of the previous Section 3 we did not mention any specific condition that assure the existence of a solution (strong or weak does not matter) of the stochastic differential equation. To prove the results with discrete time observation some more is needed.

A0. There exists a constant $K > 0$ such that

$$|S(x) - S(y)| \leq K|x - y|, \quad |\sigma(x) - \sigma(y)| \leq K|x - y|.$$

Under this condition, the SDE (1) has a unique strong solution X.

A3. The diffusion coefficient is bounded: $\sigma_*^2 := \sup_{x \in \mathbb{R}} \sigma(x)^2 < \infty$. The invariant density $f_{S,\sigma}$ is bounded.

A4. $\sup_{t \in [0,\infty)} E|X_t|^2 < \infty$.

Condition **A0**, **A3** and **A4** are essential in discrete sample case. The following condition deals with the rate of convergence of the high frequency sample scheme and the rate of convergence of the sequence $\{b_n\}$.

A5. In addition to $h_n = O(n^{-2/3}(\text{Log } n)^{1/3})$, which implies $nh_n^2 \to 0$, we assume the following:

(i) $b_n^{-2} h_n \cdot \text{Log } n \cdot \text{Log } m(n) \to 0$;

(ii) $b_n \text{Log } m(n) \to 0$.

Typically, $\text{Log } m(n) = O(\text{Log } n^\alpha)$ for some $\alpha > 0$. In this case, the above (i) and (ii) are satisfied if we take $b_n = n^{-1/4}\text{Log } n$.

The main result is the following (see Masuda et al., [19])

Theorem 3. *Assume* **A0** – **A5** *for* (S_0, σ). *Under* $H_0 : S = S_0$, *it holds that*

$$\mathcal{D}^n = \frac{\sup_{x \in [-\infty,\infty]} |U^n(x)|}{\widehat{\Sigma}^n} \to^d \sup_{t \in [0,1]} |B_t|,$$

where $\{B_t : t \geq 0\}$ *is a standard Brownian motion and* $\widehat{\Sigma}^n$ *is a consistent estimator for* $\Sigma_{S_0,\sigma}$.

The random field U^n takes values in $\ell^\infty([-\infty, \infty])$, and the random field V^n takes values in $C_\rho([-\infty, \infty])$ almost surely. Moreover we prove the uniform approximation $\sup_{x \in [-\infty,\infty]} |U^n(x) - V^n(x)| \to^p 0$ that is essential to obtain the asymptotic result. This latest result follows by the *exponential inequality* for continuous martingales, from the *maximal inequality* for general random variables (see van der Vaart and Wellner [27]) and from the following Lemma.

Lemma 1. *For every $\varepsilon > 0$ there exists a constant $K > 0$ such that*

$$\limsup_n \mathbf{P}\left(\frac{1}{Log\ n}\max_{1\le i\le n}\frac{\sup_{t\in(t_{i-1}^n,t_i^n]}|X_t - X_{t_{i-1}^n}|^2}{t_i^n - t_{i-1}^n} \ge K\right) < \varepsilon.$$

To prove it, it is sufficient to show that

$$E\left(\max_{1\le i\le n}\frac{\sup_{t\in(t_{i-1}^n,t_i^n]}|X_t - X_{t_{i-1}^n}|^2}{t_i^n - t_{i-1}^n}\right) = O(\text{Log } n).$$

A consistent estimator for $\Sigma_{S_0,\sigma}$ is given by the following

Lemma 2. *Assume* **A0** – **A4** *for (S_0, σ). The estimator*

$$\widehat{\Sigma}^n = \sqrt{\frac{1}{t_n^n}\sum_{i=1}^n |X_{t_i^n} - X_{t_{i-1}^n}|^2}$$

is consistent for $\Sigma_{S,\sigma}$.

To prove the last theorem note that it is sufficient that h_n goes to zero.

In order to prove the consistency of the test we have to study the test statistics under the alternative. We denote by $\mathcal{S}$ the class of functions S which satisfies **A0** – **A4** and

$$\int_{-\infty}^{x_S}(S(z) - S_0(z))f_{S,\sigma}(z)dz \neq 0 \quad \text{for some } x_S \in (-\infty, \infty]. \qquad (3)$$

The precise description of our problem is testing the null hypothesis H_0 : $S = S_0$ versus the alternatives $H_1 : S \in \mathcal{S}$.

The following result guarantees the consistency of the test.

Theorem 4. *Assume* **A0**–**A5** *for* (S, σ)*. Under* $H_1 : S \in \mathcal{S}$*, if* $\widehat{\Sigma}^n$ *is bounded in probability, then it holds that*

$$\mathcal{D}^n = \frac{\sup_{x \in [-\infty, \infty]} |U^n(x)|}{\widehat{\Sigma}^n} \neq O_P(1).$$

5 Tick time observations

Tick time sample scheme, roughly speaking, consists in observing the underling process only when the process reaches some fixed valued of a suitable grid in the state space. The moments when the process reaches those values are called tick times. Tick times arises in many problems in finance, when for example, the prices are sampled with every *continued price changes* in bid or ask quotation data. See for example the work of Fukasawa [10] and reference therein. Usually in finance the most common scheme is one where the prices are sampled at regular interval in calendar time. With the increasing availability of transaction data alternative sampling scheme, such as tick time sampling and transaction time sampling, has gain popolarity. See Griffin and Oomen [13] for an interesting discussion on these different sample scheme. The technique used for discrete time observation can be extended to construct a test statistics when observations are sampled as in the following

Sampling scheme. For every $T > 0$, let $\cup_p (a_p^T, a_{p+1}^T]$ be a countable partition of $(-\infty, +\infty)$ and assume that $\inf_p |a_{p+1}^T - a_p^T| > 0$ for each T. The process $X = \{X_t : t \geq 0\}$ is observed at random times $0 = \tau_0^T < \tau_1^T < \cdots < \tau_{n(T)}^T < \tau_{n(T)+1}^T = T$ where $N(T) = \sup\{i : \tau_i^T < T\}$, $\tau_0^T = 0$

$$\tau_1^T = \inf\{t > 0 : X_t = a_p^T \quad \text{for some} \quad p\}$$

and

$$\tau_i^T = \inf\{t > \tau_{i-1}^T : X_t = a_{p-1}^T \text{ or } X_t = a_{p+1}^T \text{ if } X_{\tau_{i-1}^T} = a_p^T\} \quad i \geq 2$$

We suppose that $h_T = o(T^{-1/2})$ as $T \to \infty$ where $h_T = \sup_p |a_{p+1}^T - a_p^T|$

To define the test statistics, as we did in the discrete time sample case, we introduce an array of constants in the state space if the process, $-\infty = x_0^T < x_1^T < x_2^T < \cdots < x_{m(T)}^T < x_{m(T)+1}^T = \infty$ such that, as $T \to \infty$,

$$\max_{2 \leq k \leq m(T)} |x_k^T - x_{k-1}^T| \to 0, \quad x_1^T \downarrow -\infty, \quad x_{m(T)}^T \uparrow \infty.$$

For example, one may consider $x_k^T = -[T] + (k/[T])$ with $k = 1, 2, ..., 2[T]^2$. Next we introduce a sequence of functions $\{\psi_k^T\}_{k,[T]}$ defined on $(-\infty, \infty)$ which approximates the indicator function $1_{(-\infty, x_k^T]}$.

Definition 2. *Let a sequence of positive constants* b_T *be given. For every* $k = 1, 2, ..., m(T)$*,* ψ_k^T *is the continuous, piecewise linear function on* $(-\infty, \infty)$ *defined by*

$$\psi_k^T(z) = \begin{cases} 1, & z \in (-\infty, x_k^T], \\ \textit{line}, & z \in [x_k^T, x_k^T + b_T], \\ 0, & z \in [x_k^T + b_T, \infty). \end{cases}$$

Also we define $\psi_0^T \equiv 0$ *and* $\psi_{m(T)+1}^T \equiv 1$.

This function satisfies the following properties:

$$|\psi_k^T(z) - \psi_k^T(z')| \leq b_T^{-1}|z - z'|;$$

$$|\psi_k^T(z) - 1_{(-\infty, x_k^T]}(z)| \leq 1_{[x_k^T, x_k^T + b_T]}(z).$$

Instead of condition **A3** we introduce the following condition for the pair

(S, σ) which is assumed in the tick sample case.

A3'. $\sup_{z\in\mathbb{R}}(1+|z|^2)f_{S,\sigma}(z) < \infty$

Moreover instead of condition **A5** we assume the following

A5'. In addition to $h_T = o(T^{-1/2})$, we assume the following:

(i) $b_T^{-1} h_T \cdot \sqrt{\log m(T)} \to 0$;

(ii) $b_T \log m(T) \to 0$.

Typically, $\log m(T) = O(\log T^\alpha)$ for some $\alpha > 0$. In this case, the above (i) and (ii) are satisfied if we take $b_T = T^{-1/4}$.

Under **A2**, **A3'** and **A4** we can prove that $N(T) < \infty$ almost surely. We approximate the random process V^T given by (2) by the random field $U^T = \{U^T(x) : x \in [-\infty, \infty]\}$ defined by

$$U^T(x) = \frac{1}{\sqrt{T}} \sum_{i=1}^{n} \psi_k^T(X_{\tau_{i-1}^T})[X_{\tau_i^T} - X_{\tau_{i-1}^T} - S_0(X_{\tau_{i-1}^T})|\tau_i^T - \tau_{i-1}^T|]$$

for $x \in (x_{k-1}^T, x_k^T],\ 1 \leq k \leq m(T) + 1$.

The main result is the following (see Negri and Nishiyama, [23])

Theorem 5. *Assume* **A0** – **A2**, **A3'**, **A4** *and* **A5'** *for* (S_0, σ)*. Under* $H_0 : S = S_0$*, it holds that*

$$\mathcal{T}^n = \frac{\sup_{x\in[-\infty,\infty]} |U^T(x)|}{\widehat{\Sigma}_2^T} \to^d \sup_{t\in[0,1]} |B_t|,$$

where

$$\widehat{\Sigma}_2^T = \sqrt{\frac{1}{T} \sum_{i=1}^{N(T)+1} |X_{\tau_i^T} - X_{\tau_{i-1}^T}|^2}$$

is a consistent estimator for $\Sigma_{S_0,\sigma}$ *and* $\{B_t : t \geq 0\}$ *is a standard Brownian motion.*

The proof is based on the uniform approximation $\sup_{x\in[-\infty,\infty]} |U^T(x) - V^T(x)| \to^p 0$ that is essential to obtain the asymptotic result. Also in this case, with similar argument as in the discrete case sample scheme, we obtain that the test is consistent under any alternative as in (3).

6 Conclusions

We have reviewed some results on goodness of fit test for the drift coefficient of a diffusion process when the observations comes from different sample schemes. We have proved that the proposed tests statistics are asymptotically distribution free and so we can calculate exact reject regions. For diffusion models this is, up to our knowledge, the first result in a nonparametric framework. We have also proved that the tests are consistent for any fixed alternative. One may think that, for example in the discrete sample case, would be more natural to consider the random field $\widetilde{U}^n = \{\widetilde{U}^n(x); x \in [-\infty, \infty]\}$ given by

$$\widetilde{U}^n(x) = \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^{n} 1_{(-\infty,x]}(X_{t_{i-1}^n})[X_{t_i^n} - X_{t_{i-1}^n} - S_0(X_{t_{i-1}^n})|t_i^n - t_{i-1}^n|]$$

(that is, the case $b_n = 0$) instead of U^n. This is the most natural approximation of the random process V^T given by (2). At least in our proof, the uniform approximation $\sup_{x\in[-\infty,\infty]} |U^n(x) - V^n(x)| \to^p 0$ is due to the continuity in z of the function ψ_k^n, so it does not seem easy to translate the result for V^n into that for $\widetilde{U}^n$. However, it is conjectured that the same result would hold also for $\widetilde{U}^n$; see a simulation study in Masuda *et al.* [19]. The same argument can be done for the tick time sample case.

References

[1] Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Rev. Financial Studies*, **9**, 385–426.

[2] Chen S. X., Gao J. (2008). A test for model specication of diffusion processes. *Ann. Statist.* 36, 1, 167-198.

[3] Dachian, S. and Kutoyants, Yu.A. (2008). On the goodness-of-fit 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.

[4] Dachian, S., Kutoyants, Y.A. (2009). On the Goodness-of-Fit Tests for Some Continuous Time Processes. Preprint: arXiv:0903.4642v1

[5] Delgado, M. and Stute, W. (2008). Distribution-free specification tests of conditional models. *J. Econometrics* **143**, 37-55.

[6] Durrett, R. (1996). *Stochastic Calculus: A Practical Introduction*, CRC Press, Boca Raton.

[7] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Volume II.* (2nd ed.) John Wiley & Sons, Inc., New York.

[8] Fournie, E., (1992). Un test de type Kolmogorov-Smirnov pour processus de diffusion ergodiques, *Rapporte de recherche*, **1696**, INRIA, Sophia-Antipolis.

[9] Fournie, E. and Kutoyants, Yu.A., (1993). Estimateur de la distance maximale pour le processus de diffusion ergodiques, *Rapporte de recherche*, **1952**, INRIA, Sophia-Antipolis.

[10] Fukasawa, M. (2008). Central limit theorem for the realized volatility based on tick time sampling. To appear on *Finance and Stochastics*.

[11] Gikhman, I.I., Skorohod, A.V. (1972). *Stochastic Differential Equations*, Springer–Verlag, New York.

[12] Gobet, E., Hoffmann, M. and Reiss, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32**, 2223-2253.

[13] Griffin, J.E. , Oomen, R.C.A. (2008). Sampling returns for realized variance calculations: tick time or transaction time? *Econometric Review*, 27, 1-3, p. 230–253.

[14] Lee, S. and Wee, I. (2008). Residual empirical process for diffusion processes. *J. Korean Math. Soc.* **45** 683-693.

[15] Koul, H. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* (2nd ed.) Springer, New York.

[16] Koul, H.L., Stute, W. (1999). Nonparametric model checks for times series, *The Annals of Statistics*, **27**, 1, 204-236.

[17] Kutoyants, Y.A. (2004). *Statistical Inference for Ergodic Diffusion Processes*, Springer, New York.

[18] Kutoyants, Yu.A., (2009). On the Goodness-of-Fit Testing for Ergodic Diffusion Processes, Prépublication 09-2, Université du Maine.

[19] Masuda, H., Negri, I. and Nishiyama, Y. (2008). *Goodness of t test for ergodic di?usions by discrete time observations: an innovation martingale approach.* Research Memorandum 1069, Inst. Statist. Math., Tokyo.

[20] Negri, I. (1998). Stationary distribution function estimation for ergodic diffusion process. *Statistical Inference for Stochastic Processes*, **1**, 61–84.

[21] Negri, I. and Nishiyama, Y. (2008). Goodness of fit test for ergodic diffusion processes. To appear on *Annals of the Institute of Statistical Mathematics*, Published on-line.

[22] Negri, I. and Nishiyama, Y. (2008). Goodness of fit test for small diffusions by discrete observations. To appear on *Annals of the Institute of Statistical Mathematics*,

[23] Negri, I. and Nishiyama, Y. (2009). Goodness of t test for ergodic di?usions by tick time sampling scheme. *Research Memorandum* **1096**, Inst. Statist. Math., Tokyo.

[24] Nishiyama, Y. (1999). A maximal inequality for continuous martingales and M-estimation in a Gaussian white noise model. *Ann. Statist.* **27**, 675-696.

[25] Nishiyama, Y. (2000). *Entropy Methods for Martingales.* CWI Tract **128**, Centrum voor Wiskunde en Informatica, Amsterdam.

[26] Nishiyama, Y. (2009). Nonparametric goodness of fit tests for ergodic diffusion processes by discrete observations. *Research Memorandum* **1094**, Inst. Statist. Math., Tokyo.

[27] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.

[28] van der Vaart, A.W. and van Zanten, H. (2005). Donsker theorems for diffusions: necessary and sufficient conditions. *Ann. Probab.* **33**, 1422-1451.

[29] Van Zanten H. (2003). On empirical processes for ergodic diffusions and rates of convergence of M-estimators, *Scandinavian Journal of Statistcs*, **30**, 443-458.