Ann Inst Stat Math (2011) 63:211–225
DOI 10.1007/s10463-009-0228-2

Goodness of fit test for small diffusions by discrete time observations

Ilia Negri · Yoichi Nishiyama

Received: 11 July 2008 / Revised: 11 December 2008 / Published online: 29 April 2009
© The Institute of Statistical Mathematics, Tokyo 2009

Abstract We consider a nonparametric goodness of fit test problem for the drift coefficient of one-dimensional small diffusions. Our test is based on discrete time observation of the processes, and the diffusion coefficient is a nuisance function which is "estimated" in some sense in our testing procedure. We prove that the limit distribution of our test is the supremum of the standard Brownian motion, and thus our test is asymptotically distribution free. We also show that our test is consistent under any fixed alternative.

Keywords Small diffusion process · Discrete time observations · Asymptotically distribution free test

1 Introduction

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 twentieth century, and they are *asymptotically distribution free*. See, e.g. Durbin (1973) for a review of the classical theory. On the other hand, the diffusion process models have been paid much attention because they

I. Negri
University of Bergamo, Viale Marconi 5, 24044 Dalmine (BG), Italy
e-mail: ilia.negri@unibg.it

Y. Nishiyama (✉)
The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan
e-mail: nisiyama@ism.ac.jp

are useful in many applications such as Biology, Medicine, Physics and Financial Mathematics. However, the problem of goodness of fit tests for diffusion processes has still been a new issue in recent years. Kutoyants (2004) considered this problem for ergodic diffusion models in his Sect. 5.4, but his tests are not asymptotically distribution free. Dachian and Kutoyants (2008) proposed some asymptotically distribution free tests for small diffusion models as well as ergodic diffusion models. However, all their results are based on *continuous time observation* of the diffusion processes. The main contribution of the present paper is that our test is based on *discrete time observation*, which is more realistic in applications.

Consider a one-dimensional stochastic differential equation (SDE)

$$X_t = x_0 + \int_0^t S(X_s)\mathrm{d}s + \varepsilon \int_0^t \sigma(X_s)\mathrm{d}W_s, \quad t \in [0, T], \tag{1}$$

where $x_0 \in \mathbb{R}$ is a deterministic initial value, S and σ are functions which satisfy some properties described in Sect. 2, and $t \rightsquigarrow W_t$ is a standard Wiener process defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$. Here, $T > 0$ is a fixed time. We consider a case where a unique strong solution X to this SDE exists, and we will consider the asymptotic as $\varepsilon \downarrow 0$. Statistical inference for this model based on continuous observation was studied by Kutoyants (1994). As for discrete observation cases, many researchers have treated the model in some parametric settings; see, e.g. Sørensen and Uchida (2003) and references therein. In this paper, we are interested in nonparametric 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 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 in Sect. 4.

We consider the following situation.

Sampling Scheme. The process $X = \{X_t; t \in [0, T]\}$ is observed at times $0 = t_0^\varepsilon < t_1^\varepsilon < \cdots < t_{n(\varepsilon)}^\varepsilon = T$, such that $h_\varepsilon = o(\varepsilon)$ as $\varepsilon \downarrow 0$, where $h_\varepsilon = \max_{1\le i\le n(\varepsilon)} |t_i^\varepsilon - t_{i-1}^\varepsilon|$. ♢

We may assume $\varepsilon \le 1$ and $h_\varepsilon \le 1$ without loss of generality. We will propose an asymptotically distribution free test based on this sampling scheme, namely, *high frequency data*. We should also mention that, in general settings, there is a huge literature on the discrete time approximations of statistical estimators for diffusion processes; see, e.g. the Introduction of Gobet et al. (2004) for a review including not only high frequency data but also low frequency data. However, it seems difficult (or impossible?) to obtain asymptotically distribution free results based on low frequency data, because in such a case the structure of the limit covariance depends on the true model (S, σ) in a complicated way.

The organization of the article is as follows. In Sect. 2, we state some conditions for (S, σ) which are assumed throughout this work. Section 3 gives the main result under the null hypothesis, assuming the existence of a consistent estimator for the limit variance. In Sect. 4, we prove that our test is consistent under any fixed alternatives, assuming the existence of a consistent estimator for the limit variance again. Some consistent estimators for the limit variance are explicitly constructed in Sect. 5. We present some simulation results in Sect. 6; they report that our test works well even when ε is not so small. The proofs for lemmas and a theorem in Sect. 5 will be given in Sect. 7, with help from the Appendix.

2 Preliminaries

Let us list some conditions for the pair of functions (S, σ).
A1. There exists a constant $C > 0$ such that

$$|S(x) - S(y)| \le C|x - y|, \quad |\sigma(x) - \sigma(y)| \le C|x - y|.$$

$\diamondsuit$

Under **A1**, the SDE (1) has a unique strong solution X, and notice also that there exists a constant $C' > 0$ such that

$$|S(x)| \le C'(1 + |x|), \quad |\sigma(x)| \le C'(1 + |x|).$$

To see this, just put $y = 0$. The constant C' depends on the values $S(0)$ and $\sigma(0)$, however the constant C itself depends on the choice of the functions (S, σ). So it is convenient to introduce the notation

$$K_{S,\sigma} = \max\{C, C'\}.$$

Also, since x_0 is a constant, it holds that $\sup_{t\in[0,T]} E|X_t|^k < \infty$ for any $k > 0$ (see, e.g., Theorem 4.6 of Liptser and Shiryaev (2001)).

Let us fix some more notations. For given S, let us denote by $x^S = \{x_t^S; t \in [0, T]\}$ the solution to the ordinary differential equation

$$\frac{\mathrm{d}x_t^S}{\mathrm{d}t} = S(x_t^S) \quad \text{with the initial value } x_0^S = x_0.$$

A2. $\Sigma_{S,\sigma} := \sqrt{\int_0^T \sigma(x_t^S)^2 dt} > 0$. (This is satisfied if $\inf_x \sigma(x) > 0$.) $\diamondsuit$

Let us close this section with making some conventions. We denote by $C[0, T]$ the space of continuous functions on $[0, T]$, and by $\ell^\infty[0, T]$ the space of bounded functions on $[0, T]$. We equip both the spaces with the uniform metric. We denote by "$\to^p$" and "$\to^d$" the convergence in probability and in distribution as $\varepsilon \downarrow 0$, respectively. The notation "$\to$" always means that we take the limit as $\varepsilon \downarrow 0$. We denote by 1_A the indicator function of the set A.

3 Asymptotically distribution free test

Throughout all this section, we shall suppose that **A1** and **A2** are satisfied for some (S_0, σ).

Our test statistics is based on the random process $U^\varepsilon = \{U^\varepsilon(u); u \in [0,T]\}$ defined by

$$U^\varepsilon(u) = \varepsilon^{-1} \sum_{i=1}^{n(\varepsilon)} 1_{[0,u]}(t_i^\varepsilon)\big[X_{t_i^\varepsilon} - X_{t_{i-1}^\varepsilon} - S_0(X_{t_{i-1}^\varepsilon})|t_i^\varepsilon - t_{i-1}^\varepsilon|\big].$$

We will approximate U^ε by the following random processes $V^\varepsilon = \{V^\varepsilon(u); u \in [0,T]\}$ and $M^\varepsilon = \{M_u^\varepsilon; u \in [0,T]\}$, defined respectively by:

$$V^\varepsilon(u) = \varepsilon^{-1} \sum_{i=1}^{n(\varepsilon)} 1_{[0,u]}(t_i^\varepsilon) \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} [\mathrm{d}X_s - S_0(X_s)\mathrm{d}s];$$

$$M_u^\varepsilon = \varepsilon^{-1} \int_0^u [\mathrm{d}X_s - S_0(X_s)\mathrm{d}s].$$

Lemma 1 $\sup_{u\in[0,T]} |U^\varepsilon(u) - V^\varepsilon(u)| \to^p 0.$

Lemma 2 $\sup_{u\in[0,T]} |V^\varepsilon(u) - M_u^\varepsilon| \to^p 0.$

Lemma 3 *$M^\varepsilon \to^d G$ in $C[0,T]$, where $G = \{G(u); u \in [0,T]\}$ is a Brownian motion with co-variance*

$$EG(u)G(v) = \int_0^{u\wedge v} \sigma(x_t^{S_0})^2 \mathrm{d}t.$$

Combining these lemmas, we obtain the following result.

Theorem 4 *$U^n \to^d G$ in $\ell^\infty[0,T]$, where G is the process appearing in Lemma 3.*

By the continuous mapping theorem, we have the following.

Corollary 5 *It holds that*

$$\sup_{u\in[0,T]} |U^\varepsilon(u)| \to^d \sup_{t\in[0,\Sigma_{S_0,\sigma}^2]} |B_t| =^d \Sigma_{S_0,\sigma} \sup_{t\in[0,1]} |B_t|,$$

where $t \rightsquigarrow B_t$ is a standard Brownian motion, and the notation "$=^d$" means that the distributions are the same.

So we have the main result of the paper.

Theorem 6 *Under $\mathrm{H}_0 : S = S_0$, suppose that $\widehat{\Sigma}^\varepsilon$ is a positive consistent estimator for $\Sigma_{S_0,\sigma}$. Then we have*

$$D^\varepsilon = \frac{\sup_{u\in[0,T]} |U^\varepsilon(u)|}{\widehat{\Sigma}^\varepsilon} \to^d \sup_{t\in[0,1]} |B_t|,$$

where $t \rightsquigarrow B_t$ is a standard Brownian motion.

The construction of a consistent estimator $\widehat{\Sigma}^{\varepsilon}$ for $\Sigma_{S,\sigma}$ will be discussed in Sect. 5. It is well known that the distribution function of the limit is given by

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

see e.g. page 343 of Feller (1971).

4 Consistency of the test

Let S_0 be that in Sect. 3. We denote by $\mathcal{S}$ the class of functions S satisfying **A1**, **A2** and

$$\int_0^{u_S} \left(S(x_t^S) - S_0(x_t^S)\right)\mathrm{d}t \neq 0 \quad \text{for some } u_S \in [0,T]. \qquad (3)$$

The precise description of our problem is testing the null hypothesis $\mathrm{H}_0 : S = S_0$ versus the alternatives $\mathrm{H}_1 : S \in \mathcal{S}$.

We will prove that our test is consistent. Fix $S \in \mathcal{S}$. We can write $U^{\varepsilon} = U_S^{\varepsilon} + U_{\Delta}^{\varepsilon}$ where

$$U_S^{\varepsilon}(u) = \varepsilon^{-1}\sum_{i=1}^{n(\varepsilon)} 1_{[0,u]}(t_i)\left[X_{t_i^{\varepsilon}} - X_{t_{i-1}^{\varepsilon}} - S(X_{t_{i-1}^{\varepsilon}})|t_i^{\varepsilon} - t_{i-1}^{\varepsilon}|\right]$$

and

$$U_{\Delta}^{\varepsilon}(u) = \varepsilon^{-1}\sum_{i=1}^{n(\varepsilon)} 1_{[0,u]}(t_i^{\varepsilon})\left(S(X_{t_{i-1}^{\varepsilon}}) - S_0(X_{t_{i-1}^{\varepsilon}})\right)|t_i^{\varepsilon} - t_{i-1}^{\varepsilon}|.$$

Now we have

$$\sup_{u\in[0,T]} |U^{\varepsilon}(u)| \ge \sup_{u\in[0,T]} |U_{\Delta}^{\varepsilon}(u)| - \sup_{u\in[0,T]} |U_S^{\varepsilon}(u)|.$$

Since S satisfies **A1** and **A2**, by the same argument as in Sect. 3, the random process U_S^{ε} converges to the corresponding Gaussian random process with S_0 replaced by S. So the second term of the right hand side is $O_P(1)$. As for the first term of the right hand side, we have the following claim.

Lemma 7 *Choose $u_S \in [0,T]$ as in (3). Then it holds that $|U_{\Delta}^{\varepsilon}(u_S)| \neq O_P(1)$.*

We therefore obtain the consistency of the test.

Theorem 8 *Under* $\mathrm{H}_1 : S \in \mathcal{S}$, *if* $\widehat{\Sigma}^\varepsilon$ *is bounded in probability, then*

$$D^\varepsilon = \frac{\sup_{u\in[0,T]} |U^\varepsilon(u)|}{\widehat{\Sigma}^\varepsilon} \neq O_P(1).$$

5 Consistent estimator for $\Sigma_{S,\sigma}$

In order to construct an asymptotically distribution free test, it is sufficient to have a consistent estimator for $\Sigma_{S,\sigma}$. The following result gives us an answer.

Theorem 9 *Suppose that* $h_\varepsilon = o(\varepsilon^2)$ *as* $\varepsilon \downarrow 0$. *For any* (S, σ) *which satisfies* **A1** *and* **A2***,*

$$\widehat{\Sigma}^\varepsilon = \sqrt{\varepsilon^{-2} \sum_{i=1}^{n(\varepsilon)} \left|X_{t_i^\varepsilon} - X_{t_{i-1}^\varepsilon}\right|^2} \tag{4}$$

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

It should be noted that when we use this estimator the constants ε on the denominator and the numerator of the test statistics D^ε are cancelled. So we can compute D^ε without knowing the value of ε.

On the other hand, the assumption $h_\varepsilon = o(\varepsilon^2)$ is stronger than the rate $h_\varepsilon = o(\varepsilon)$ assumed in Sampling Scheme. If one can construct a consistent estimator assuming only $h_\varepsilon = o(\varepsilon)$ (or if σ is assumed to be known), then the main result holds for the rate $h_\varepsilon = o(\varepsilon)$.

The following argument is due to an anonymous referee. Define

$$\xi^\varepsilon(S) = \sqrt{\varepsilon^{-2} \sum_{i=1}^{n(\varepsilon)} \left|X_{t_i^\varepsilon} - X_{t_{i-1}^\varepsilon} - S(X_{t_{i-1}^\varepsilon})|t_i^\varepsilon - t_{i-1}^\varepsilon|\right|^2}. \tag{5}$$

Then the following claims hold.

(i) Assume $h_\varepsilon = o(\varepsilon)$. If the true drift is S, then $\xi^\varepsilon(S) \to^p \Sigma_{S,\sigma}$.
(ii) Assume $h_\varepsilon = O(\varepsilon^2)$. If the true drift is S, then $\xi^\varepsilon(S) - \xi^\varepsilon(S') = O_P(1)$.

The proofs are not difficult. [For (i) use also Itô's formula. To see (ii) observe that $|(\sum_i |x_i|^p)^{1/p} - (\sum_i |y_i|^p)^{1/p}| \leq (\sum_i |x_i - y_i|^p)^{1/p}$ by Minkowski's inequality. So $|\xi^\varepsilon(S) - \xi^\varepsilon(S')|^2 \leq \varepsilon^{-2} \sum_{i=1}^{n(\varepsilon)} \{|S(X_{t_{i-1}^\varepsilon}) - S'(X_{t_{i-1}^\varepsilon})||t_i^\varepsilon - t_{i-1}^\varepsilon|\}^2$ and the expectation of the right hand side is $O(\varepsilon^{-2} h_\varepsilon)$.] Using these facts, we may set $\widehat{\Sigma}^\varepsilon = \xi^\varepsilon(S_0)$; then it is consistent for $\Sigma_{S_0,\sigma}$ under $H_0 : S = S_0$ and bounded in probability under $H_1 : S \in \mathcal{S}$. The sampling scheme $h_\varepsilon = O(\varepsilon^2)$ is sufficient, and this enables us to treat the case $h_\varepsilon = t_i^\varepsilon - t_{i-1}^\varepsilon = T/n$ and $\varepsilon = n^{-1/2}$.

6 Simulation studies

In this section, for simplicity we consider the equidistant sampling case that is $h_\varepsilon = t_i^\varepsilon - t_{i-1}^\varepsilon = \frac{1}{n}$ for every $i \le n$, where n is the number of observation in every trajectory. In practical analysis the asymptotics of Theorems 6, 8 and 9 may not be realized, due to lack of enough data (small n) or too noise (ε not so small). For this reason, we run a simulation experiment to verify the performance of the proposed test for moderate ε and n. In our experimental design we consider $n = 100, 200, 1000$ and $\varepsilon = 0.5, 0.2, 0.1, 0.05$, i.e. such that ε and $1/(n\varepsilon^2)$ are not always small.

We consider the following stochastic differential equation

$$\mathrm{d}X_t = S(X_t)\mathrm{d}t + \varepsilon \mathrm{d}W_t, \quad 0 < t \le 1, \quad X_0 = 1$$

and the following three hypotheses

$\mathcal{H}_0 : S(x) = S_0(x) = -2x,$
$\mathcal{H}_A : S(x) = S_A(x) = -4x,$
$\mathcal{H}_B : S(x) = S_B(x) = 1 - 2x.$

We study the asymptotic behavior of the test statistics D^ε with $S = S_0$. We denote by $D_0^{\varepsilon,l}$, $D_A^{\varepsilon,l}$ and $D_B^{\varepsilon,l}$ the value of the test statistics when the data are generated according to $\mathcal{H}_0$, $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, with $1 \le l \le m$ and $m = 1000$ the number of simulated trajectories for each model. It should be expected that the model specified by $\mathcal{H}_A$ is more difficult to discriminate than the model in $\mathcal{H}_B$.

We take the significance level to be $\alpha = 0.05$. If we denote by F the distribution function of the limit distribution (2), we see that $F(x) = 0.95$ when $x \doteqdot 2.24$ hence the critical region is $\{x > 2.24\}$. If the asymptotic conditions are realized $P(D^\varepsilon > 2.24)$ tends to 0.05 under $\mathcal{H}_0$ and tends to 1 under $\mathcal{H}_A$ or $\mathcal{H}_B$. We exhibit that the test may work also for moderate size of ε by our simulation study.

We compute

- the empirical size (e.s.) defined by $\#\{l : D_0^{\varepsilon,l} > 2.24\}/m$, the sampling proportion of making incorrect rejections of the null;
- the empirical power (e.p. A) defined by $\#\{l : D_A^{\varepsilon,l} > 2.24\}/m$, the sampling proportion of making successful rejection of the null, when the data comes from model (A).
- the empirical power (e.p. B) defined by $\#\{l : D_B^{\varepsilon,l} > 2.24\}/m$, the sampling proportion of making successful rejection of the null, when the data comes from model (B).

Table 1 reports the empirical size and the empirical power for the two different alternatives $\mathcal{H}_A$ and $\mathcal{H}_B$ when $\widehat{\Sigma}^\varepsilon$ given by (4) is used as the estimator for $\Sigma_{S_0,\sigma}$. The simulation results show that for moderate ε, i.e. $\varepsilon = 0.50$ and 0.20, the test behaves as the asymptotic expects because the asymptotic on $1/(n\varepsilon^2)$ is realized.

Table 2 reports the case where $\widehat{\Sigma}^\varepsilon = \xi^\varepsilon(S_0)$ given by (5) is used as the estimator for $\Sigma_{S_0,\sigma}$. The simulation results are better than the preceding case, due to the fact that the asymptotic condition $1/(n\varepsilon^2) = O(1)$ is realized.

Table 1 Empirical size (e.s.) when the true model is specified by $\mathcal{H}_0$, and empirical power (e.p. A) and (e.p. B) when the true model is specified, respectively, by model (A) and model (B), based on m trajectories for different values of ε and n, where the estimator for $\Sigma_{S_0,\sigma}$ is the one given by (4)

	ε			
	0.50	0.20	0.10	0.05
$n = 100$				
e.s.	0.036	0.028	0.005	0.000
e.p. A	0.035	0.526	0.996	1.000
e.p. B	0.472	0.996	1.000	1.000
$1/(n\varepsilon^2)$	(0.040)	(0.250)	(1.000)	(4.000)
$n = 200$				
e.s.	0.044	0.044	0.014	0.000
e.p. A	0.045	0.728	1.000	1.000
e.p. B	0.477	0.997	1.000	1.000
$1/(n\varepsilon^2)$	(0.020)	(0.125)	(0.500)	(2.000)
$n = 1000$				
e.s	0.047	0.053	0.045	0.020
e.p. A	0.059	0.858	1.000	1.000
e.p. B	0.484	0.997	1.000	1.000
$1/(n\varepsilon^2)$	(0.004)	(0.025)	(0.100)	(0.400)

In parenthesis the value of $1/(n\varepsilon^2)$ is reported

7 Proofs

Proof of Lemma 1 Without loss of generality, we may assume that $\varepsilon \leq 1$ and $h_\varepsilon \leq 1$. It follows from Lemma 12 that

$$
\begin{aligned}
E\left(\sup_{u\in[0,T]}\left|U^\varepsilon(u) - V^\varepsilon(u)\right|\right) &\leq \varepsilon^{-1} E \sum_{i=1}^{n(\varepsilon)} \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} \left|S_0(X_{t_{i-1}^\varepsilon}) - S_0(X_s)\right| \mathrm{d}s \\
&\leq \varepsilon^{-1} \sum_{i=1}^{n(\varepsilon)} \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} K_{S_0,\sigma} E\left|X_{t_{i-1}} - X_s\right| \mathrm{d}s \\
&\leq \varepsilon^{-1} T K_{S_0,\sigma} C_1 \max\left\{h_\varepsilon, \varepsilon h_\varepsilon^{1/2}\right\} \\
&\to 0.
\end{aligned}
$$

So we have the assertion of the lemma. □

Proof of Lemma 2 Notice that

$$
M_u^\varepsilon = \varepsilon^{-1} \sum_{i=1}^{n} \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} 1_{[0,u]}(s)\,[\mathrm{d}X_s - S_0(X_s)\mathrm{d}s]
$$

Table 2 Empirical size (e.s.) when the true model is specified by $\mathcal{H}_0$, and empirical power (e.p. A) and (e.p. B) when the true model is specified, respectively, by model (A) and model (B), based on m trajectories for different values of ε and n, where the estimator for $\Sigma_{S_0,\sigma}$ is $\widehat{\Sigma}^\varepsilon = \xi^\varepsilon(S_0)$ given by (5)

	ε			
	0.50	0.20	0.10	0.05
$n = 10$				
e.s.	0.022	0.038	0.068	0.221
e.p. A	0.019	0.121	0.283	0.338
e.p. B	0.333	0.998	1.000	1.000
$1/(n\varepsilon^2)$	(0.400)	(2.500)	(10.00)	(40.00)
$n = 100$				
e.s.	0.039	0.053	0.047	0.052
e.p. A	0.051	0.812	1.000	1.000
e.p. B	0.459	0.997	1.000	1.000
$1/(n\varepsilon^2)$	(0.040)	(0.250)	(1.000)	(4.000)
$n = 200$				
e.s.	0.042	0.057	0.048	0.051
e.p. A	0.056	0.851	1.000	1.000
e.p. B	0.470	0.997	1.000	1.000
$1/(n\varepsilon^2)$	(0.020)	(0.125)	(0.500)	(2.000)
$n = 1000$				
e.s	0.047	0.055	0.049	0.051
e.p. A	0.061	0.870	1.000	1.000
e.p. B	0.483	0.997	1.000	1.000
$1/(n\varepsilon^2)$	(0.004)	(0.025)	(0.100)	(0.400)

In parenthesis the value of $1/(n\varepsilon^2)$ is reported

$$
\begin{aligned}
&= V^\varepsilon(u) + \varepsilon^{-1} \int_{t_{i-1}^\varepsilon}^u [\mathrm{d}X_s - S_0(X_s)\mathrm{d}s] \quad \forall u \in [t_{i-1}^\varepsilon, t_i^\varepsilon) \\
&= V^\varepsilon(u) + \int_{t_{i-1}^\varepsilon}^u \sigma(X_s)\mathrm{d}W_s \quad \forall u \in [t_{i-1}^\varepsilon, t_i^\varepsilon)
\end{aligned}
$$

and that $M_T^\varepsilon = V^\varepsilon(T)$. Now we have

$$
\begin{aligned}
E\left|\sup_{u\in[0,T]} |V^\varepsilon(u) - M_u^\varepsilon|\right|^4 &\le \sum_{i=1}^{n(\varepsilon)} E \sup_{u\in[t_{i-1}^\varepsilon, t_i^\varepsilon)} |V^\varepsilon(u) - M_u^\varepsilon|^4 \\
&\le \sum_{i=1}^{n(\varepsilon)} E \sup_{u\in[t_{i-1}^\varepsilon, t_i^\varepsilon]} \left|\int_{t_{i-1}^\varepsilon}^u \sigma(X_s)\mathrm{d}W_s\right|^4 .
\end{aligned}
$$

It follows from Burkholder–Davis–Gundy's inequality (see e.g. Theorem 26.12 of Kallenberg 2002) that, for a constant c_k depending only on $k = 4$, the right hand side is bounded by

$$
\begin{aligned}
c_4 \sum_{i=1}^{n(\varepsilon)} E \left| \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} \sigma(X_s) \mathrm{d}s \right|^2 &\leq c_4 \sum_{i=1}^{n(\varepsilon)} E \left(\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} 1 \mathrm{d}s \cdot \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} \sigma(X_s)^2 \mathrm{d}s \right) \\
&\leq c_4 T \max_{1 \leq i \leq n(\varepsilon)} \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} E\sigma(X_s)^2 \mathrm{d}s \\
&\leq c_4 T h_\varepsilon \sup_{s \in [0,T]} E\sigma(X_s)^2 \\
&\to 0.
\end{aligned}
$$

The proof is finished. $\square$

Proof of Lemma 3 When $S = S_0$, it holds that

$$
M_u^\varepsilon = \int_0^u \sigma(X_s) \mathrm{d}W_s
$$

We will apply the central limit theorem for continuous martingales.

$$
\begin{aligned}
\langle M^\varepsilon \rangle_u &= \int_0^u \sigma(X_s)^2 \mathrm{d}s \\
&= \int_0^u \left(\sigma(X_s)^2 - \sigma(x_s^{S_0})^2 \right) \mathrm{d}s + \int_0^u \sigma(x_s^{S_0})^2 \mathrm{d}s \\
&= (I) + (II).
\end{aligned}
$$

Now, using Lemma 10, we have

$$
\begin{aligned}
|(I)| &\leq \int_0^u \left| \sigma(X_s)^2 - \sigma(x_s^{S_0})^2 \right| \mathrm{d}s \\
&\leq \int_0^T \left| \sigma(X_s) - \sigma(x_s^{S_0}) \right| \left| \sigma(X_s) + \sigma(x_s^{S_0}) \right| \mathrm{d}s \\
&\leq K_{S_0,\sigma} \sup_{t \in [0,T]} \left| X_t - x_t^{S_0} \right| \cdot \int_0^T \left| \sigma(X_s) + \sigma(x_s^{S_0}) \right| \mathrm{d}s \\
&\leq K_{S_0,\sigma} \exp(K_{S_0,\sigma} T) \cdot \varepsilon \sup_{t \in [0,T]} \left| \int_0^t \sigma(X_s) \mathrm{d}W_s \right| \cdot \int_0^T \left| \sigma(X_s) + \sigma(x_s^{S_0}) \right| \mathrm{d}s \\
&= O_P(\varepsilon).
\end{aligned}
$$

So we have $\langle M^\varepsilon \rangle_u \to^p \int_0^u \sigma(x_s^{S_0})^2 \mathrm{d}s$, and the weak convergence of the process $u \rightsquigarrow M_u^\varepsilon$ holds. $\square$

Proof of Lemma 7 We simply denote $u = u_S$. We consider the following random variables:

$$\begin{aligned}
A_1^\varepsilon &= \varepsilon U_\Delta^\varepsilon(u) \\
&= \sum_{i=1}^{n(\varepsilon)} 1_{[0,u]}(t_i^\varepsilon) \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} \big(S(X_{t_{i-1}^\varepsilon}) - S_0(X_{t_{i-1}^\varepsilon})\big) \mathrm{d}s; \\
A_2^\varepsilon &= \sum_{i=1}^{n(\varepsilon)} 1_{[0,u]}(t_i^\varepsilon) \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} (S(X_s) - S_0(X_s)) \mathrm{d}s; \\
A_3^\varepsilon &= \sum_{i=1}^{n(\varepsilon)} \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} 1_{[0,u]}(s)(S(X_s) - S_0(X_s)) \mathrm{d}s; \\
A_4 &= \int_0^u (S(x_s^S) - S_0(x_s^S)) \mathrm{d}s.
\end{aligned}$$

First, it holds that

$$\begin{aligned}
E|A_1^\varepsilon - A_2^\varepsilon| &\leq \sum_{i=1}^{n(\varepsilon)} E \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} \left\{ \left| S(X_{t_{i-1}^\varepsilon}) - S(X_s) \right| + \left| S_0(X_{t_{i-1}^\varepsilon}) - S_0(X_s) \right| \right\} \mathrm{d}s \\
&\leq \left(K_{S,\sigma} + K_{S_0,\sigma} \right) \sum_{i=1}^{n(\varepsilon)} E \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} |X_{t_{i-1}^\varepsilon} - X_s| \mathrm{d}s \\
&\leq \left(K_{S,\sigma} + K_{S_0,\sigma} \right) T C_1 h_\varepsilon^{1/2} \\
&\to 0,
\end{aligned}$$

where C_1 is a constant appearing in Lemma 12. So we have $|A_1^\varepsilon - A_2^\varepsilon| \to^p 0$.

Next,

$$\begin{aligned}
A_2^\varepsilon - A_3^\varepsilon &= \sum_{i=1}^{n(\varepsilon)} \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} \left(1_{[0,u]}(t_i^\varepsilon) - 1_{[0,u]}(s) \right) \{ S(X_s) - S_0(X_s) \} \, \mathrm{d}s \\
&= \int_{t_{i-1}^\varepsilon}^{u} \{ S(X_s) - S_0(X_s) \} \, \mathrm{d}s \quad \forall u \in [t_{i-1}^\varepsilon, t_i^\varepsilon).
\end{aligned}$$

If $u = u_S = T$, then $A_2^\varepsilon = A_3^\varepsilon$. Since

$$\max_{1 \leq i \leq n(\varepsilon)} \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} \{ |S(X_s)| + |S_0(X_s)| \} \, \mathrm{d}s \leq h_\varepsilon \sup_{s \in [0,T]} \{ |S(X_s)| + |S_0(X_s)| \} = O_P(h_\varepsilon),$$

we have $|A_2^\varepsilon - A_3^\varepsilon| \to^p 0$.

Finally, notice that

$$A_3^\varepsilon - A_4 = \int_0^u (S(X_s) - S_0(X_s))\,\mathrm{d}s - \int_0^u \left(S(x_s^S) - S_0(x_s^S)\right)\mathrm{d}s.$$

It follows from Lemma 10 that

$$\begin{aligned}
\int_0^u |S(X_s) - S(x_s^S)|\mathrm{d}s &\leq \int_0^T K_{S,\sigma}|X_s - x_s^S|\mathrm{d}s \\
&\leq T K_{S,\sigma} \sup_{t\in[0,T]} |X_t - x_t^S| \\
&\leq T K_{S,\sigma} \cdot \exp(K_{S,\sigma}T) \cdot \varepsilon \sup_{t\in[0,T]} \left|\int_0^t \sigma(X_s)\mathrm{d}W_s\right| \\
&= O_P(\varepsilon).
\end{aligned}$$

By the same way, it holds that

$$\begin{aligned}
\int_0^u \left|S_0(X_s) - S_0(x_s^S)\right| \mathrm{d}s &\leq T K_{S_0,\sigma} \cdot \exp(K_{S,\sigma}T) \cdot \varepsilon \sup_{t\in[0,T]} \left|\int_0^t \sigma(X_s)\mathrm{d}W_s\right| \\
&= O_P(\varepsilon).
\end{aligned}$$

Thus we have $|A_3^\varepsilon - A_4| \to 0$.

Consequently, we obtain $A_1^\varepsilon \to^p A_4 \neq 0$, which implies that $|U_\Delta^\varepsilon(u_S)| \neq O_P(1)$. □

Proof of Theorem 9 By Itô's formula, we have

$$|X_{t_i^\varepsilon}|^2 - |X_{t_{i-1}^\varepsilon}|^2 = 2\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} X_s \mathrm{d}X_s + \varepsilon^2 \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} \sigma(X_s)^2 \mathrm{d}s.$$

Since

$$|\widehat{\Sigma}^\varepsilon|^2 = \varepsilon^{-2} \sum_{i=1}^{n(\varepsilon)} \left\{ |X_{t_i^\varepsilon}|^2 - |X_{t_{i-1}^\varepsilon}|^2 - 2X_{t_{i-1}^\varepsilon}(X_{t_i^\varepsilon} - X_{t_{i-1}^\varepsilon}) \right\},$$

it is enough to show that

$$\varepsilon^{-2} \sum_{i=1}^{n(\varepsilon)} \int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon} (X_s - X_{t_{i-1}^\varepsilon})\mathrm{d}X_s \to^p 0$$

and

$$\int_0^T \sigma(X_s)^2 \mathrm{d}s \to^p \Sigma_{S,\sigma}^2.$$

The latter is proved by the same argument as that in the proof of Lemma 3. As for the former, observe that

$$\varepsilon^{-2}\left|\sum_{i=1}^{n(\varepsilon)}\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon}(X_s-X_{t_{i-1}^\varepsilon})\mathrm{d}X_s\right|$$
$$\leq \varepsilon^{-2}\sum_{i=1}^{n(\varepsilon)}\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon}|X_s-X_{t_{i-1}^\varepsilon}||S(X_s)|\mathrm{d}s+\varepsilon^{-1}\left|\sum_{i=1}^{n(\varepsilon)}\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon}(X_s-X_{t_{i-1}^\varepsilon})\sigma(X_s)\mathrm{d}W_s\right|$$

By Lemma 12, the expectation of the first term on the right hand side is

$$\begin{aligned}
&\varepsilon^{-2}\sum_{i=1}^{n(\varepsilon)}\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon}E\left(|X_s-X_{t_{i-1}^\varepsilon}||S(X_s)|\right)\mathrm{d}s\\
&\leq \varepsilon^{-2}\sum_{i=1}^{n(\varepsilon)}\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon}\sqrt{E|X_s-X_{t_{i-1}^\varepsilon}|^2}\sqrt{E|S(X_s)|^2}\mathrm{d}s\\
&\leq \varepsilon^{-2}\sum_{i=1}^{n(\varepsilon)}\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon}\sqrt{C_2\max\{h_\varepsilon^2,\varepsilon^2h_\varepsilon\}}\sqrt{E|S(X_s)|^2}\mathrm{d}s\\
&\leq \varepsilon^{-2}T\sqrt{C_2}\max\left\{h_\varepsilon,\varepsilon h_\varepsilon^{1/2}\right\}\cdot\sup_{s\in[0,T]}\sqrt{E|S(X_s)|^2}\\
&\to 0.
\end{aligned}$$

On the other hand, the expectation of the square of the second term on the right hand side is

$$\begin{aligned}
&\varepsilon^{-2}\sum_{i=1}^{n(\varepsilon)}E\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon}|X_s-X_{t_{i-1}^\varepsilon}|^2\sigma(X_s)^2\mathrm{d}s\\
&\leq \varepsilon^{-2}\sum_{i=1}^{n(\varepsilon)}\int_{t_{i-1}^\varepsilon}^{t_i^\varepsilon}\sqrt{E|X_s-X_{t_{i-1}^\varepsilon}|^4}\sqrt{E\sigma(X_s)^4}\mathrm{d}s\\
&\leq \varepsilon^{-2}T\sqrt{C_4h_\varepsilon^2}\sup_{s\in[0,T]}\sqrt{E\sigma(X_s)^4}\\
&\to 0.
\end{aligned}$$

This proves the consistency of our estimator. □

Appendix

In the main part of this article, we use the following lemmas all of which are trivial or well-known.

Lemma 10 *For any solution $X = \{X_t; t \in [0, T]\}$ to the SDE (1), it holds that*

$$\sup_{t\in[0,T]} |X_t - x_t| \leq \exp(K_{S,\sigma} T) \cdot \varepsilon \sup_{t\in[0,T]} \left| \int_0^t \sigma(X_s) \mathrm{d}W_s \right|.$$

Proof. The proof is a simple application of Gronwall's inequality: apply Lemma 4.13 of Liptser and Shiryaev (2001) for $c_0 = \varepsilon \sup_{t\in[0,T]} |\int_0^t \sigma(X_s) dW_s|$, $c_1 = K_{S,\sigma}$, $c_2 = 0$, $u(t) = |X_t - x_t|$, and $v(t) = 1$. □

Lemma 11 *Let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $|f(x)| \leq H(1 + |x|)$ for some $H > 0$. Let $X = \{X_t; t \in [0, T]\}$ be any stochastic process. Let $k > 0$ and assume $\sup_{t\in[0,T]} E|X_t|^k < \infty$. Then, it holds that $\sup_{t\in[0,T]} E|f(X_t)|^k < \infty$.*

Proof. Since $|x + y|^k \leq ||x| + |y||^k \leq |2 \max\{|x|, |y|\}|^k \leq 2^k\{|x|^k + |y|^k\}$, the lemma is trivial. □

Lemma 12 *Let $X = \{X_t; t \in [0, T]\}$ be a solution to the SDE (1) for (S, σ) which satisfies* **A1**. *For any $k > 0$, there exists a constant $C_k > 0$ such that for any $0 \leq t \leq t' \leq T$ and any $\varepsilon > 0$*

$$E|X_{t'} - X_t|^k \leq C_k \max\left\{|t' - t|^k, \varepsilon^k |t' - t|^{k/2}\right\}.$$

In particular, if $|t' - t| \leq 1$ and $\varepsilon \leq 1$, then

$$E|X_{t'} - X_t|^k \leq C_k |t' - t|^{k/2}.$$

Remark. The constant C_k is *not* a universal constant depending only on k. It actually depends on S, σ, T and the value $\sup_{t\in[0,T]} E|X_t|^k$. However, it does not depend on t, t', ε.

Proof. Use Hölder's inequality and Burkholder–Davis–Gundy's inequality. □

Acknowledgments Part of this work was performed during the first author's visit to Tokyo in December 2007 and the second author's visit to Bergamo in May 2008 supported by MIUR 2006 Grant, and the kind hospitality of the Institute of Statistical Mathematics and the University of Bergamo is greatly appreciated. The authors thank Prof. Hiroki Masuda for helpful discussion, and the anonymous referee for constructive comments, especially for the suggestion on the estimator for $\Sigma_{S_0,\sigma}$ as we stated in Sect. 5.

References

Dachian, S., Kutoyants, Yu. A. (2008). On the goodness-of-fit tests for some continuous time processes. In F. Vonta, M. Nikulin, N. Limnios, & C. Huber-Carol (Eds.), *Statistical models and methods for biomedical and technical systems* (pp. 395–413). Boston: Birkhuser.

Durbin, J. (1973). *Distribution theory for tests based on the sample distribution function.* Philadelphia: Society for Industrial and Applied Mathematics.

Feller, W. (1971). *An introduction to probability theory and its applications* (2nd ed.). New York: John Wiley & Sons, Inc.

Gobet, E., Hoffmann, M., Reiss, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *The Annals of Statistics, 32*, 2223–2253.

Kallenberg, O. (2002). *Foundations of modern probability* (2nd ed.). Heidelberg: Springer.

Kutoyants, Yu. A. (1994). *Identification of dynamical systems with small noise.* Dordrecht: Kluwer Academic Publishers.

Kutoyants, Yu. A. (2004). *Statistical inference for ergodic diffusion processes.* London: Springer.

Liptser, R. S., Shiryaev, A. N. (2001). *Statistics of random processes. I. General theory* (2nd ed.). Berlin: Springer.

Sørensen, M., Uchida, M. (2003). Small-diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli, 9*, 1051–1069.