Statistical Inference for Stochastic Processes
https://doi.org/10.1007/s11203-024-09309-2

A model specification test for nonlinear stochastic diffusions with delay

Zongwu Cai[1] · Hongwei Mei[2] · Rui Wang[1]

Received: 23 March 2023 / Accepted: 31 January 2024

Abstract

This paper investigates model specification problems for nonlinear stochastic differential equations with delay (SDDE). Compared to the model specification for conventional stochastic diffusions without delay, the observed sequence does not admit a Markovian structure so that the classical testing procedures may not be applicable. To overcome this difficulty, a moment estimator is newly proposed based on the ergodicity of SDDEs and its asymptotic properties are established. Based on the proposed moment estimator, a testing procedure is proposed for our model specification testing problems. Particularly, the limiting distributions of the proposed test statistic are derived under null hypotheses and the test power is examined under some specific alternative hypotheses. Finally, a Monte Carlo simulation is conducted to illustrate the finite sample performance of the proposed test.

Keywords Model specification test · Stochastic differential equation with delay · Moment estimator · Ergodicity · Invariant measure · Non-Markovian property

Mathematics Subject Classification Primary 62F05; Secondary 62M07

We thank the Editor (Professor Arnaud Gloter), the Associate Editor, and the anonymous referee for their constructive and helpful comments and suggestions that improved significantly the quality of the paper. This work is partially supported by Simons Foundation's Travel Support for Mathematicians Program under Grant Number 00002835 and the Natural Science Foundation of China under Grant Numbers 72133002 and 72033008.

✉ Hongwei Mei
hongwei.mei@ttu.edu

Zongwu Cai
caiz@ku.edu

Rui Wang
rui.wang@ku.edu

1 Department of Economics, The University of Kansas, Lawrence, KS 66045, USA

2 Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, USA

Published online: 22 March 2024

1 Introduction

Consider a d-dimensional stochastic differential equation with delay

$$dX(t) = b_0(X_t)dt + \sigma_0(X_t)dW(t), \tag{1}$$

where $X(t)$ denotes the state of the system at time t and $X_t = \{X(t+s) : -\tau \leq s \leq 0\}$ is called the segment process, which includes the all information of $X(\cdot)$ on $[t-\tau, t]$. Here, $\tau > 0$ is a fixed constant representing the delay structure. Two coefficient functions $b_0(\cdot)$ (drift function) and $\sigma_0(\cdot)$ (diffusion function) are appropriate mappings of the segment process and $W(t)$ is a r-dimensional standard Brownian motion with $r \geq 1$. Our interest is in testing the joint parametric family $\mathcal{P} = \{(b(\cdot;\theta), \sigma(\cdot;\theta)) : \theta \in \Theta\}$, where Θ is a compact subset of $\mathbb{R}^{m_\theta}$ with $m_\theta \geq 1$. The parametric family $\mathcal{P}$ provides explanatory power for understanding the underlying dynamics. This is to say that our aim is to test if the following null hypothesis holds or not

$$H_0 : b_0(\cdot) = b(\cdot;\theta),\ \ \sigma_0(\cdot) = \sigma(\cdot;\theta) \text{ for some } \theta \in \Theta.$$

Throughout the paper, we always write the true parameter $\theta = \theta^* \in \Theta$ if H_0 is true even though the value of θ^* may not be given. This test is about to see if a parametric (linear) model is appropriate for a real application.

When $b_0(X_t) = b_0(X(t))$ and $\sigma_0(X_t) = \sigma_0(X(t))$ for some appropriate functions $b_0(\cdot)$ and $\sigma_0(\cdot)$ in $\mathbb{R}^d$, the SDDE model in (1) reduces to a classical stochastic differential equation (SDE) without delay. The model specification testing problem for such special case has been a very important topic in the literature since the pioneer work by Aït-Sahalia (1996). For example, there are some extensions to the methods as in Aït-Sahalia (1996); Hong and Li (2005); Chen et al. (2008); Aït-Sahalia et al. (2009), especially, see Hong and Li (2005) for the kernel estimation for transition density, Chen et al. (2008) for transitional density using the empirical likelihood, and (Aït-Sahalia et al. 2009) for a specification test for the transition density of a discretely sampled continuous-time jump-diffusion process. The similar goodness-of-fit testing problems for continuous-time stochastic diffusions also receive extensive attention in the literature. For example, one may refer to the papers by Dachian and Kutoyants (2008), Negri and Nishiyama (2009), Kleptsyna and Kutoyants (2014) and the review paper by López-Pŕez et al. (2022), and references therein.

Different from the aforementioned papers, it is assumed that the joint parametric family $\mathcal{P}$ admits a delay dependence structure in our paper. The motivation of delay dependence stems from the fact that many of the phenomena witnessed in applications do not have an immediate effect from the moment of their occurrence. With such an important feature, SDDEs are widely used in stochastic modeling in practice. For example, applied works focusing on SDDEs in the literature, include, to name just a few, the works by Mao (2007), Bratsun et al. (2005), Hobson and Rogers (1998), Steiner et al. (2017), Marschak (1971), Lawrence (2012), Lei and Mackey (2007), Rihan (2021), Stoica (2005), Karatzas (1996), Hale and Lunel (2013), Chen and Yu (2014), Ivanov and Swishchuk (2008), Arriojas et al. (2007), and references therein, with particular applications in the analysis of stability in automatic control in stochastic systems, gene regulation, inertia and delay in decision-making, stochastic volatility, stochastic games, economics of information systems, optimal control in economics, and a delayed Black-Schole formulation and option pricing in finance.

The parameter estimation and statistical inference for SDDEs also receive a lot of attention in the literature, see, for example, Benke and Pap (2017), Gushchin and Küchler (1999), Küchler and Kutoyants (2000), Küchler and Sørensen (2010, 2013), Reiss (2005), and references therein. In the literature, it is commonly assumed that the drift coefficient is linear and the diffusion coefficient is constant, and the observations are in real-time in the aforementioned papers. For a different small perturbation approach, the reader is referred to the paper by Kutoyants (2021) and references therein. To the best of our knowledge, there is no work yet concerning the model specification problem for general nonlinear SDDEs especially with discrete-time observations. The paper aims to fill this gap by providing an efficient testing procedure under a general setting.

More specifically, a testing procedure is constructed based on the ergodicity of non-linear SDDEs for the model specification test. Such a generalization allows one to work on more complex model specification problems with delay in practice. Due to the non-Markovian structure, the classical testing methods using transition probability for Markovian observations as in Aït-Sahalia (1996), Hong and Li (2005), Chen et al. (2008), Aït-Sahalia et al. (2009), can not be directly applicable here. To this end, we propose a new approach, which consists of two steps. First, a moment estimator is introduced and then, its asymptotic properties are investigated. Indeed, the proposed estimator is inspired by the ergodicity of SDDEs, similar to that in Küchler and Sørensen (2013) but different from the small perturbation approach as in Kutoyants (2021). Then, based on the proposed moment estimator, a statistic is constructed and its limiting distributions are established, which can be used in the model specification test problem for SDDEs. It is worth mentioning that instead of the non-parametric Cramer Von Mises statistics, our test statistics are constructed through introducing a new function $f_0(\cdot)$ satisfying some conditions, combined with the stochastic generator of the SDDE. The key reason for this is that the invariant measure of the segment process of the SDDE is infinite-dimensional (see Assumption 2). Moreover, because the diffusion coefficient can be estimated non-parametrically using in-fill asymptotics, our methods are designed particularly for testing the drift coefficient $b_0(\cdot)$. Therefore, it is assumed that $\sigma_0(\cdot)$ is independent of θ throughout the paper.

The well-posed results (such as the existence and uniqueness) for the SDDE in (1) can be found in Section 5.2 in Mao (2007). Define an operator on $\mathcal{A}$ for any twice continuously differentiable function $f(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}$ by

$$\mathcal{A}f(\eta;\theta) = \langle b(\eta;\theta), \nabla f(\eta(0))\rangle + \frac{1}{2}\text{trace}\left[\sigma(\eta)\sigma^\top(\eta)D^2 f(\eta(0))\right],$$

where η denotes a possible path of the segment process (see (3) below), ∇f is the gradient of f, and $D^2 f$ is the Hessian matrix of f. We also write $\mathcal{A}_0 f(\eta) = \mathcal{A}f(\eta;\theta^*)$ for the true $\theta = \theta^*$. Through the well-known Itô's formula, it follows that for a regular function $f(\cdot)$ (see Section 5.6 in Mao (2007), for example), the following process

$$f(X(t)) - f(X(0)) - \int_0^t \mathcal{A}_0 f(X_s)ds$$

is a local martingale. Actually, $\mathcal{A}$ can be seen as the infinitesimal generator for the segment process $\{X_t\}$. To work on the testing problem in our paper, it is assumed the solution process to be exponential ergodic with a unique invariant measure μ. In such case, the observation is asymptotically stable which coincides with the classical stable assumptions for observations in the literature. The results concerning the exponential ergodicity can be found in the Appendix.

The rest of the paper is arranged as follows. We present the definition of moment estimator and prove its limit theorems in Sect. 2. Then, a testing procedure for testing our model specification problem is developed in Sect. 2 too. Some simulation results to justify our theory are illustrated in Sect. 3. We summarize our conclusions in Sect. 4. The mathematical proofs of the main results are relegated to Sect. 5. Finally in Appendix, some limit theorems for SDDEs are recalled, especially on the exponential ergodicity theory.

2 Specification test

2.1 Moment estimator

In this section, our aim is to present the definition of our estimator which is called a moment estimator since the definition depends on H_0 and takes a moment estimator form.

Suppose that the SDDE in (1) is observed with a time-window Δ and a sequence of observations $\{Z_i\}_{i=0}^n$ are observed, where $Z_i = X(i\Delta)$, and that there also exists a set of regular functions $\mathbf{f} = \{f_k : k = 1, \cdots, m\}$ with $m \geq m_\theta$. To emphasize the dependence of θ, we write by $\mu(\cdot;\theta)$ the unique invariant measure of X_t (see Appendix). It is well-known that for $k = 1, \cdots, m$,

$$\int_{\mathscr{C}} \mathcal{A} f_k(\eta;\theta^*)\mu(d\eta;\theta^*) = 0.$$

As X is exponential ergodic, by the law of large numbers (LLN), it is easy to obtain

$$\frac{1}{T}\int_\tau^T \mathcal{A} f_k(X_t;\theta^*)dt \to \int_{\mathscr{C}} \mathcal{A} f_k(\eta;\theta^*)\mu(d\eta;\theta^*) = 0$$

almost surely as $T \to \infty$, where $\mathscr{C}$ is defined in (3) later. Replacing the continuous-time process X above by the sequence of discrete time observations $\{Z_i\}_{i=0}^n$, define

$$\begin{aligned}&\widehat{A}_{n,\Delta}(f_k;\theta)\\ &= \frac{1}{n}\sum_{i=\ell_\Delta}^{n}\left[\tilde{b}_\Delta^\top(Z_{i-\ell_\Delta},\cdots,Z_i;\theta)\nabla f_k(Z_i) + \frac{1}{2}\mathrm{trace}\Big([\tilde{\sigma}_\Delta^\top\tilde{\sigma}_\Delta](Z_{i-\ell_\Delta},\cdots,Z_i)D^2 f_k(Z_i)\Big)\right],\end{aligned}$$

where $\ell_\Delta = \lfloor \tau/\Delta \rfloor$, the largest integer smaller than or equal to τ/Δ, and $\tilde{b}_\Delta(\cdot)$ and $\tilde{\sigma}_\Delta(\cdot)$ are some appropriate approximations chosen for $b(\cdot)$ and $\sigma(\cdot)$ in (1). Here, note that different from $b(\cdot)$ and $\sigma(\cdot)$, $\tilde{b}_\Delta(\cdot)$ and $\tilde{\sigma}_\Delta(\cdot)$ are finitely dimensional functions. For such case, $\hat{A}_{n,\Delta}(f_k;\theta)$ is essentially an approximation of

$$A_{n\Delta}(f_k;\theta) = \frac{1}{n\Delta}\int_\tau^{n\Delta}\left[b^\top(X_t;\theta)\nabla f_k(X(t)) + \frac{1}{2}\mathrm{trace}\Big([\sigma^\top\sigma](X_t)D^2 f_k(X(t))\Big)\right]dt.$$

A natural idea of defining the moment estimator is to solve the following equations for θ,

$$\widehat{A}_{n,\Delta}(f_k;\theta) = 0,\ k = 1, \cdots, m.$$

Because the solution may not exist, define the moment estimator $\hat{\theta}_{n,\Delta}(\mathbf{f})$ as an m-dimensional vector in the compact set Θ by

$$\hat{\theta}_{n,\Delta}(\mathbf{f}) = \mathrm{argmin}_{\theta\in\Theta}\sum_{k=1}^{m}\left|\widehat{A}_{n,\Delta}(f_k;\theta)\right|, \tag{2}$$

if $m \geq m_\theta$. If $m = m_\theta$ (the model is just identified), since Θ is compact, $\hat{\theta}_{n,\Delta}(\mathbf{f})$ is well-defined and is the zero-point (when n is large and Δ is small). When $m > m_\theta$, $\hat{\theta}_{n,\Delta}(\mathbf{f})$ is over-identified so that it is a minimum point only while is not the anticipated zero-point. Therefore, it needs a different effort to consider this over-identification case, which is left as a future research topic. In this paper, we consider the only case that $m = m_\theta$. Now, write the error term (after an appropriate scaling) by

$$\gamma_{n,\Delta} = \sqrt{n\Delta}\sum_{k=1}^{m}\left|\widehat{A}_{n,\Delta}(f_k;\hat{\theta}_{n,\Delta}(\mathbf{f}))\right| = \sqrt{n\Delta}\inf_{\theta\in\Theta}\sum_{k=1}^{m}\left|\widehat{A}_{n,\Delta}(f_k;\theta)\right|,$$

and then, the limit behavior of the error $\gamma_{n,\Delta}$ plays an important role in the later analytic study.

It is assumed that $\Delta = \Delta_n \to 0$ and $n\Delta_n \to \infty$ and $n \to \infty$. The condition $n\Delta_n \to \infty$ allows one to apply the LLN and the central limit theorem (CLT) for SDDEs and is also sufficient for $(n - \ell_{\Delta_n})/n \to 1$. Here, one should emphasize that the assumption $\Delta_n \to 0$ is to guarantee our moment estimator being (asymptotically) unbiased. The unbiasedness of the moment estimator is critical in our testing procedure. Because our final results are in a weak-convergence manner, it only needs to show that the CLT holds for any subsequence (n_j, Δ_{n_j}) of (n, Δ_n) with $\sum_{j=1}^{\infty}\Delta_{n_j} < \infty$. Therefore, it is assumed throughout the paper that

$$n_j \to \infty,\ \Delta_{n_j} \to 0,\ n_j\Delta_{n_j} \to \infty \text{ and } \sum_{j=1}^{\infty}\Delta_{n_j} < \infty.$$

For the sake of convenience, the subscript j is omitted in the sequel without confusion.

Until now, we have unveiled the definition of our moment estimator $\hat{\theta}_{n,\Delta_n}(\mathbf{f})$ given in (2). Our main goal is to study the consistency and asymptotic normality so that one can construct a test statistic for the testing problem.

2.2 Asymptotic properties

First, we establish the consistency of our moment estimator under H_0 with the detailed proof, relying on the ergodicity theory for SDDEs, given in the Appendix. To this end, define the functional space

$$\mathscr{C} = \Big\{\eta : [-\tau, 0] \mapsto \mathbb{R}^d | \eta(\cdot) \text{ is continuous on } [-\tau, 0]\Big\}, \tag{3}$$

equipped with the sup-norm metric $\|\eta\|_{\mathscr{C}} = \sup_{-\tau\le s\le 0}|\eta(s)|$. For any $\eta(\cdot) \in \mathscr{C}$, we write the δ-increment functional by

$$w_\delta(\eta) = \sup_{\substack{-\tau\le u\le v\le 0\\ |u-v|\le\delta}}|\eta(u) - \eta(v)|.$$

Now, the following two assumptions are needed for investigating the large sample theory. The first assumption is from Bao et al. (2020) to guarantee the ergodicity of SDDE.

Assumption 1 (A1). $\sigma(\cdot)$ is Lipschitz continuous; $b(\cdot) : \mathscr{C} \times \Theta \mapsto \mathbb{R}^d$ and $\sigma(\cdot) : \mathscr{C} \times \Theta \mapsto \mathbb{R}^{d\times r}$ is continuous, and bounded on bounded subsets of $\mathscr{C}$.

(A2). There exist two constants $\lambda_1, \lambda_2 > 0$ such that $\lambda_1 > \lambda_2 e^{-\lambda_1\tau}$ such that

$$2\langle\xi(0) - \eta(0), b(\xi;\theta^*) - b(\eta;\theta^*)\rangle \le -\lambda_1|\xi(0) - \eta(0)|^2 + \lambda_2\|\xi - \eta\|_{\mathscr{C}}^2.$$

(A3). $\sigma\sigma^\top(\cdot,\theta^*)$ is invertible with

$$\sup_{\mathscr{C}}\Big(|\sigma\sigma^\top(\cdot,\theta^*)|+|(\sigma\sigma^\top)^{-1}(\cdot,\theta^*)|\Big)<\infty,$$

and $|\partial_\theta b(\eta;\theta)|\leq L(1+\|\eta\|_{\mathscr{C}}),\ |\partial_\theta\sigma(\eta;\theta)|\leq L$, where L is a tentative constant which may vary from place to place, in what follows.

The second assumption has two perspectives: the identifiable condition and the approximation of discrete-time observations to the true solution of SDDE.

Assumption 2 Suppose that $f_k(\cdot)\in\mathbf{f}$ is twice continuously differentiable with bounded second order derivatives satisfying:
(i) θ^* is the unique solution

$$\int_{\mathscr{C}}\mathcal{A}f_k(\eta;\theta)\mu(d\eta;\theta^*)=0 \text{ for all } k=1,\cdots,m,$$

where μ is the unique invariant measure of X_t.
(ii) The rank of the matrix $R(\theta^*)=(r_1(\theta^*),\cdots,r_m(\theta^*))$ is m, where

$$r_k(\theta)=\int_{\mathscr{C}}\partial_\theta\mathcal{A}f_k(\eta;\theta)\mu(d\eta;\theta). \tag{4}$$

(iii) For any $\eta\in\mathscr{C}$, there exist $\tilde b_\Delta:(\mathbb{R}^d)^{\ell_\Delta+1}\times\Theta\mapsto\mathbb{R}^d$ and $\tilde\sigma_\Delta:(\mathbb{R}^d)^{\ell_\Delta+1}\times\Theta\mapsto\mathbb{R}^{d\times r}$ such that

$$\begin{aligned}
&|\tilde b_\Delta\big(\eta(-\ell_\Delta*\Delta),\cdots,\eta(0);\theta\big)|+|\tilde\sigma_\Delta\big(\eta(-\ell_\Delta*\Delta),\cdots,\eta(0)\big)|\leq L(1+\theta)(1+\|\eta\|_{\mathscr{C}}),\\
&|\partial_\theta b(\eta;\theta)|+|\partial_\theta^2 b(\eta;\theta)|+|\partial_\theta\tilde b_\Delta(\eta(-\ell_\Delta*\Delta),\cdots,\eta(0);\theta)|\\
&\qquad+|\partial_\theta^2\tilde b_\Delta(\eta(-\ell_\Delta*\Delta),\cdots,\eta(0);\theta)|\leq L(1+\|\eta\|_{\mathscr{C}}),\\
&|b(\eta;\theta)-\tilde b_\Delta\big(\eta(-\ell_\Delta*\Delta),\cdots,\eta(0);\theta\big)|+|\sigma(\eta;\theta)-\tilde\sigma_\Delta\big(\eta(-\ell_\Delta*\Delta),\cdots,\eta(0);\theta\big)|\\
&\quad\leq L(1+|\theta|)w_\Delta(\eta),
\end{aligned}$$

and

$$|\partial_\theta b(\eta;\theta)-\partial_\theta\tilde b_\Delta(\eta(-\ell_\Delta*\Delta),\cdots,\eta(0);\theta)|\leq Lw_\Delta(\eta).$$

To prove the consistency, the following proposition is needed.

Proposition 1 *Suppose Assumptions 1 and 2 hold. It follows that*

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\widehat{A}_{n,\Delta_n}(f_k;\theta)-\int_{\mathscr{C}}\mathcal{A}f_k(\eta;\theta)\mu(d\eta;\theta^*)\right|=0$$

almost surely.

Now, let us present the consistency of the estimator under H_0 with its proof given in Sect. 5.

Theorem 1 *(Consistency) Suppose Assumptions 1 and 2 hold. Under H_0, it follows that*

$$\hat\theta_{n,\Delta_n}\to\theta^* \tag{5}$$

almost surely as $n\to\infty$. Consequently, the moment estimator is consistent.

Next, we establish the asymptotic normality for the moment estimator $\hat{\theta}_{n,\Delta_n}(\mathbf{f})$ defined in (2). To this end, we proceed with the following estimate for the error term γ_{n,Δ_n} with its proof given in Sect. 5.

Lemma 1 *Under Assumptions 1 and 2, if Θ has neighborhood of θ^*, then $\gamma_{n,\Delta_n} = 0$ almost surely as n goes to infinity.*

From the estimate for γ_{n,Δ_n} in Lemma 1, our testing problem is considered by two different cases (I): Θ has a neighborhood of θ^*; (II): $\Theta = \{\theta^*\}$. It is obvious that Case II is trivial in the estimation step so that our focus is only on Case I. The following is our result on the asymptotic normality of $\hat{\theta}_{n\Delta_n}(\mathbf{f})$ with its proof given in Sect. 5.

Theorem 2 *(Asymptotic Normality) Suppose Assumptions 1 and 2 hold and Θ has a neighborhood of θ^*. As $n \to \infty$ and $\sqrt{n}\Delta_n \to 0$, then, it follows that under H_0,*

$$\sqrt{n\Delta_n}\big[\hat{\theta}_{n,\Delta_n}(\mathbf{f}) - \theta^*\big] \to N\left(0, \Sigma(\mathbf{f};\theta^*)\right) \tag{6}$$

in distribution, where $\Sigma(\mathbf{f};\theta^*)$ is defined as

$$r_k^\top(\theta^*)\Sigma(\mathbf{f};\theta^*)r_k(\theta^*) = \int_{\mathscr{C}} |\sigma^\top(\eta)\nabla f_k(\eta(0))|^2\mu(d\eta;\theta^*) \text{ for all } k = 1,\cdots,m \tag{7}$$

with $r_k(\theta)$ defined in (4).

Note that the true value of θ^* is not obtainable, we provide the following asymptotic normality with variance being independent of θ^*. Together with Lemma 2 in the Appendix, we have the following proposition.

Proposition 2 *Suppose assumptions in Lemma 2 in Appendix and Assumption 2 hold. As $n \to \infty$ and $\sqrt{n}\Delta_n \to 0$, it follows that under H_0,*

$$\sqrt{n\Delta_n}\cdot\Sigma^{-1/2}(\mathbf{f};\hat{\theta}_{n,\Delta_n}(\mathbf{f}))\left[\hat{\theta}_{n,\Delta_n}(\mathbf{f}) - \theta^*\right] \to N(0,1)$$

in distribution, where $\Sigma(\mathbf{f};\theta^*)$ is defined in (7).

Until now, we have obtained the asymptotic normality for our moment estimator $\hat{\theta}_{n,\Delta}(\mathbf{f})$. While we cannot directly apply such asymptotic normality to our hypothesis testing problem as the true θ^* is not obtainable. Therefore, we need to construct a statistic for our model testing problem in Sect. 2.3, described in the next section.

2.3 Test statistic

In this section, a statistic is constructed. To this end, let $f_0(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}$ and define the statistic as follows:

$$\begin{aligned}\hat{A}_{n,\Delta}(f_0;\hat{\theta}_{n,\Delta}(\mathbf{f})) = \sqrt{\frac{\Delta}{n}}\sum_{i=\ell_\Delta}^{n}\Big[&\tilde{b}_\Delta^\top(Z_{i-\ell_\Delta},\cdots,Z_i;\hat{\theta}_{n,\Delta}(\mathbf{f}))\nabla f_0(Z_i)\\ &+\frac{1}{2}\mathrm{trace}\Big([\tilde{\sigma}_\Delta^\top\tilde{\sigma}_\Delta](Z_{i-\ell_\Delta},\cdots,Z_i)D^2 f_0(Z_i)\Big)\Big].\end{aligned}$$

Now, it is ready to present the main results of the paper. The first theorem concerns the asymptotic normality for Case I in Theorem 3 and the second theorem is for Case II in Theorem 4 with their detailed proofs given in Sect. 5.

Theorem 3 *Let all assumptions in Proposition 2 hold. Suppose that $f_0(\cdot)$ is twice continuously differentiable with bounded second-order derivatives. Under H_0, if $\sigma(f_0, \mathbf{f}; \theta^*) \neq 0$, as $n \to \infty$ and $\sqrt{n}\Delta_n \to 0$, one has*

$$\hat{T}_{n,\Delta_n}(f_0, \mathbf{f}; \hat{\theta}_{n,\Delta_n}(\mathbf{f})) = \sigma^{-1}(f_0, \mathbf{f}; \hat{\theta}_{n,\Delta_n}(\mathbf{f})) \cdot \hat{A}_{n,\Delta_n}(f_0; \hat{\theta}_{n,\Delta_n}(\mathbf{f})) \to N(0,1) \tag{8}$$

in distribution, where $r_0(\theta) = \int_{\mathscr{C}} \partial_\theta \mathcal{A} f_0(\eta;\theta)\mu(d\eta;\theta)$ and

$$\sigma^2(f_0, \mathbf{f}; \theta) = \int_{\mathscr{C}} \left[\left[R^{-1}(\theta) r_0(\theta) \right]^\top \begin{pmatrix} \sigma^\top(\eta)\nabla f_1(\eta(0)) \\ \vdots \\ \sigma^\top(\eta)\nabla f_m(\eta(0)) \end{pmatrix} - \sigma^\top(\eta)\nabla f_0(\eta(0)) \right]^2 \mu(d\eta;\theta).$$

Here, $f_0(\cdot)$ can not be a linear combination of $\mathbf{f}$ because $\sigma^2(f_0, \mathbf{f}; \theta) = 0$ for such a case. This coincides with our intuition from the definition of our moment estimator as an infimum point using $\mathbf{f}$.

Theorem 4 *Let $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ with $\theta_0 \neq \theta_1$. Under H_0, as $n \to \infty$ and $\sqrt{n}\Delta_n \to 0$ if $v^{-1}(\mathcal{A} f_0(\cdot;\theta_0);\theta_0) \neq 0$, one has*

$$v^{-1}(\mathcal{A} f_0(\cdot;\theta_0);\theta_0) \cdot \hat{A}_{n,\Delta_n}(f_0;\theta_0) \to N(0,1) \tag{9}$$

in distribution, where $v(\cdot)$ is defined in (19) in Appendix. Moreover, under H_1, if $f_0(\cdot)$ satisfies $\int_{\mathscr{C}} \mathcal{A} f_0(\eta;\theta_0)\mu(d\eta;\theta_1) \neq 0$, then,

$$v^{-1}(\mathcal{A} f_0(\cdot;\theta_0);\theta_0) \cdot \hat{A}_{n,\Delta_n}(f_0;\theta_0) \to \infty$$

in probability so that the test power converges to 1 under H_1 with a rate of $(n\Delta_n)^{-1}$.

Remark 1 First, from Theorems 3 and 4, one can conclude that the probabilities of falsely rejecting H_0 for both Case I and Case II are asymptotically α as $n \to \infty$, $\Delta_n \to 0$, $n\Delta_n \to \infty$ and $\sqrt{n}\Delta_n \to 0$. Second, even though the closed forms of $\sigma(f_0, \mathbf{f}; \theta)$ and $v(\mathcal{A} f(\cdot;\theta);\theta)$ as functions of θ may not be obtainable, their values can be computed numerically through an independent Monte-Carlo method without using the observations. Therefore, $\sigma(f_0, \mathbf{f}; \theta)$ and $v(\mathcal{A} f(\cdot;\theta);\theta)$ are treated as known functions in testing procedure. Finally, it seems that selecting $f_0(\cdot)$ is important, so that $\int_{\mathscr{C}} \mathcal{A} f_0(\eta;\theta^*)\mu(d\eta;\theta_1) \neq 0$ is for a good test power, which is illustrated by a concrete example in simulation study later (see the results given in Table 4 later).

With the above limiting results, the proposed testing procedure can be summarized as follows:
Case I (i.e. Θ has a neighborhood of θ^*): reject H_0 if $|\sigma^{-1}(f_0, \mathbf{f}; \hat{\theta}_{n,\Delta_n}(\mathbf{f})) \cdot \hat{A}_{n,\Delta_n}(f_0, \hat{\theta}_{n,\Delta_n}(\mathbf{f}))| \geq z_{\alpha/2}$, where $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile of a standard normal distribution.
Case II (i.e $H_0 : \theta = \theta^*$): reject H_0 if $|v^{-1}(\mathcal{A} f_0(\cdot;\theta^*);\theta^*) \cdot \hat{A}_{n,\Delta_n}(f_0, \theta^*)| \geq z_{\alpha/2}$.

3 Monte Carlo simulation study

We consider testing a generalized Vasicek model as in Vasicek (1977) with delay,

$$dX(t) = [a_0 - b_0 X(t) + \theta b_1(X(t-\tau))]dt + \sigma dW(t)$$

Table 1 The test sizes for different significance levels α and number of observations n with $\Delta = 10^{-3}$ and $\theta^* = 1$

α	0.01	0.05	0.10
$n = 10^4$	0.006	0.030	0.064
$n = 10^6$	0.012	0.054	0.110

for some $\theta \in \Theta = [-l, l]$. Since our main focus is on the delay structure in our paper, set a_0, b_0, and σ as given constants in our simulation example. $b_1(\cdot)$ is a non-linear function, which distinguishes our results from the previous result in Küchler and Sørensen (2013) and so on.

Since θ is one dimensional in our example, we pick only a function $f(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ such that $\int_{\mathscr{C}} f'(\eta(0))\eta(-\tau)\mu(d\eta; \theta^*) \neq 0$. Our moment estimator is defined by

$$\hat{\theta}_{n,\Delta} = \Pi \left[\frac{\sum_{i=\ell_\Delta}^{n} [(b_0 Z_i - a_0) f'(Z_i) - \sigma^2 f''(Z_i)/2]}{\sum_{i=\ell_\Delta}^{n} [b_1(Z_{i-\ell_\Delta}) f'(Z_i)]} \right],$$

where Π is the projection from $\mathbb{R}$ to $[-l, l]$. By (20) in the Appendix, a simple calculation yields that

$$r(\theta) = \int_{\mathscr{C}} f'(\eta(0)) b_1(\eta(-\tau))\mu(d\eta; \theta)(\neq 0) \text{ and } r_0(\theta) = \int_{\mathscr{C}} f_0'(\eta(0)) b_1(\eta(-\tau))\mu(d\eta; \theta).$$

Also, it is not difficult to see that

$$\sigma^2(f_0, \mathbf{f}; \theta) = \sigma^2 \int_{\mathscr{C}} \left[f_0'(\eta(0)) - \frac{r_0(\theta)}{r(\theta)} f'(\eta(0)) \right]^2 \mu(d\eta; \theta).$$

Now, let us present a concrete example to illustrate our results. Set $a_0 = 0$, $b_0 = 5$, $\sigma = 1$, $\tau = 0.1$, $b_1(x) = I(|x| < 1)$, $f(x) = x$, and $f_0(x) = x^2/2$. For this case the moment estimator and testing statistics are

$$\begin{aligned} &\hat{T}_{n,\Delta}(f_0, \mathbf{f}; \hat{\theta}_{n,\Delta}(\mathbf{f})) \\ &\quad = \sigma^{-1}(f_0, \mathbf{f}; \hat{\theta}_{n,\Delta}(\mathbf{f})) \sqrt{\frac{\Delta}{n}} \sum_{i=\ell_\Delta}^{n} \left(Z_i[a_0 - b_0 Z_i + \hat{\theta}_{n,\Delta}(\mathbf{f}) b_1(Z_{i-\ell_\Delta})] + 1/2 \right), \end{aligned}$$

where

$$\hat{\theta}_{n,\Delta}(\mathbf{f}) = \Pi \left[\frac{\sum_{i=\ell_\Delta}^{n} [b_0 Z_i - a_0]}{\sum_{i=\ell_\Delta}^{n} b_1(Z_{i-\ell_\Delta})} \right], \quad \sigma^2(f_0, \mathbf{f}; \theta) = \sigma^2 \int_{\mathscr{C}} \left[\eta(0) - \frac{r_0(\theta)}{r(\theta)} \right]^2 \mu(d\eta; \theta),$$

and Π is the projection from $\mathbb{R}$ to Θ. Here, as mentioned earlier, $\sigma^2(f_0, \mathbf{f}; \cdot)$ can be calculated by an independent Monte-Carlo simulation without using the observations. When simulating the observations, set the step size $\delta = \Delta/10$ and $n_\delta = 10 * n$ recursions to obtain $\{Y_i\}_{i=0}^{n_\delta}$ for the SDDE, where $Y_i = X(i\delta)$. Then, the observations are $\{Z_i\}_{i=0}^{n}$, where $Z_i = Y_{i*\Delta/\delta}$. The simulation is repeated 500 times.

Table 1 reports the test sizes for different values of sample size n. From Table 1, one can see clearly that the test size converges to the nominal sizes, when the sample size n becomes large (proportional to the observation window Δ).

In Table 2, the test powers are listed if the alternative hypothesis takes $H_1 : b(\eta) = a_0 - b_0\eta(0) + \theta^* b_1(\eta(-0.1))$ with $a_0 \neq 0$ (i.e. $\theta_1 = \theta^*$). Such alternatives correspond to the cases when the perturbation of H_1 from H_0 is a constant in the drift coefficient. When a_0 departures from 0, the test power tends to one quickly. This means that indeed, the proposed test is powerful.

Table 2 The test powers for different values of a_0 in $H_1 : b(\eta;\theta) = a_0 - b_0\eta(0) + \theta_1 b_1(\eta(-0.1))$ with $\theta_1 = 1$, $\alpha = 0.05$, $n = 10^6$ and $\Delta = 10^{-3}$

$\alpha = 0.05$							
a_0	0.3	0.2	0.1	0	−0.1	−0.2	−0.3
Power	1.000	0.998	0.588	0.048	0.454	0.898	0.982

Table 3 The test powers for different values of θ_1 under $H_0 : b(\eta;\theta) = -b_0\eta(0) + \theta_0 b_1(\eta(-0.1))$ versus $H_1 : b(\eta;\theta) = -b_0\eta(0) + \theta_1 b_1(\eta(-0.1))$ with $\alpha = 0.05$, $n = 10^6$ and $\Delta = 10^{-3}$

$\alpha = 0.05$							
θ_1	0.7	0.8	0.9	1.0	1.1	1.2	1.3
Power	0.936	0.770	0.286	0.048	0.460	0.962	1.000

Table 4 The test powers for different values of θ_0 and a_0 under $H_0 : b(\eta;\theta) = a_0 - b_0\eta(0) + \theta_0 b_1(\eta(-0.1))$ versus $H_1 : b(\eta;\theta) = a_0 - b_0\eta(0)$ with $\alpha = 0.05$, $n = 10^6$ and $\Delta = 10^{-3}$

$a_0 = 1$ θ_0	0	0.3	0.5	1
Power	0.056	0.9960	1.000	1.000
$a_0 = 0$				
θ_0	0	0.3	0.5	1
Power	0.044	0.048	0.044	0.046

In Table 3, the hypothesis is set as $H_0 : b(\eta;\theta) = a_0 - b_0\eta(0) + \theta_0 b_1(\eta(-0.1))$ versus $H_1 : b(\eta;\theta) = a_0 - b_0\eta(0) + \theta_1 b_1(\eta(-0.1))$, where H_1 is indexed by θ_1. The new feature for such an example is that $\gamma_{n,\Delta} \to 0$ in probability fails. For this case, $\hat{\theta}_{n,\Delta}(\mathbf{f}) = 1$ and the testing statistic becomes

$$
\begin{aligned}
&\Big(\int_{\mathscr{C}} [\eta(0)]^2 \mu(d\eta;\theta^*)\Big)^{-1/2} \hat{A}_{n,\Delta}(f_0;\theta^*) \\
&\quad = \Big(\int_{\mathscr{C}} [\eta(0)]^2 \mu(d\eta;\theta^*)\Big)^{-1/2} \sqrt{\frac{\Delta}{n}} \sum_{i=\ell_\Delta}^{n} \Big([a_0 - b_0 Z_i + \theta_0 b_1(Z_{i-\ell_\Delta} < 1)]Z_i + 1/2\Big).
\end{aligned}
$$

Note that the true value of $\int_{\mathscr{C}} [\eta(0)]^2 \mu(d\eta;\theta^*)$ is 0.1392 by performing an independent Monte-Carlo simulation in prior. Table 3 summarizes the test powers for this case. From Table 3, one also can observe that when θ_1 departures from 1, the test power tends to one quickly, which implies that the proposed test works reasonably well.

Finally, Table 4 displays the test powers for different θ_0 and a_0 if the stochastic diffusion admits no delay structure in H_1. When $a_0 = 1$, we can see that the power tends to 1 very quickly when θ_0 departs from the true value 0. This concludes our test is very powerful in distinguishing the delay structure from conventional stochastic diffusions. While if $a_0 = 0$, our tests are not becoming more powerful when θ_0 departs from 0. The reason is that $\int_{\mathscr{C}} \mathcal{A} f_0(\eta;\theta_0)\mu(d\eta;\theta_1) = 0$ in such a case, which justifies our third conclusion in Remark 1. To make our tests powerful, a different $f_0(\cdot)$ rather than $f_0(x) = x^2/2$ should be selected.

4 Conclusion

In this paper, a model specification test for SDDEs is proposed based on its ergodicity. Compared to model specification problems for stochastic diffusions without delay, the observation does not admit a Markovian structure. The proposed method allows us to work with the case that the stochastic diffusions have nonlinear coefficients and admits a delay structure under the null hypothesis. Through Monte Carlo simulation, we observe that the proposed test has a good test size and is indeed powerful.

Before concluding the paper, we would like to discuss how to apply our method if the observed window Δ is fixed. Such a case for linear SDDEs with additive diffusions was studied in Küchler and Sørensen (2013). Due to the special structure assumed there, it is asserted that the conditional distribution of $X_{(i+1)\Delta}$ on $X_\Delta, \cdots, X_{i\Delta}$ is normal which plays an essential role in their study. Otherwise, a biased estimator can be concluded in Küchler and Sørensen (2011). Because the diffusions are assumed non-linear in our problem, such a property fails and their method is not applicable here.

In our paper, the key of selecting $\mathcal{A}f_k(\cdot)$ as moment functions lies in the fact that $\widehat{A}_{n,\Delta}(f_k;\theta^*)$ is asymptotic to 0 (independent of θ^*). While for fixed Δ, the limit of $\widehat{A}_{n,\Delta}(f_k;\theta^*)$ depends on θ^* and therefore, the moment functions $\{\mathcal{A}f_k(\cdot)\}$ would not be appropriate in this case. To propose an appropriate moment estimator for such case, it needs to find $g_\Delta(\eta;\theta)$ such that $g_\Delta(\eta;\theta)$ depends on the observable part in η only and $\int_{\mathscr{C}} g_\Delta(\eta;\theta^*\mu(d\eta;\theta^*) = 0$. The choice is not easy in general because the explicit form of the invariant measure for the segment process is not obtainable. This problem is left as a future study. To summarize, in this paper, we let $\Delta \to 0$, which leads to the closed forms of mean and variance in the asymptotic normality. Our problem can be seen as a model specification testing problem for non-linear SDDEs with high-frequency data. Finally, as aforementioned, it would be very interesting to investigate the over-identified case ($m > m_\theta$) issue and the efficiency of the proposed moment estimator should be explored too, which are left future research topics.

5 Mathematical proofs

Note that $O_p(1)$ stands for a term which is bounded in probability and $o_p(1)$ means that it converges to 0 in probability.

Proof of Proposition 1 Note that

$$\mathbb{E}\Big|\widehat{A}_{n,\Delta}(f_k;\theta)-(n\Delta)^{-1}\int_\tau^{n\Delta}\mathcal{A}f_k(X_t;\theta)dt\Big|^2 \le Ln^{-1}\mathbb{E}\sum_{i=1}^n(1+\|X_{i\Delta}\|_{\mathscr{C}})^2\cdot\|w_\Delta(X_{i\Delta})^2\|$$

$$\le Ln^{-1}\sum_{i=1}^n\sqrt{\mathbb{E}(1+\|X_{i\Delta}\|_{\mathscr{C}})^4\cdot\mathbb{E}\|w_\Delta(X_{i\Delta})^4\|} = L\Delta.$$

Since $\sum_{j=1}^\infty \Delta_{n_j} < \infty$, then,

$$\sum_{j=1}^\infty\mathbb{E}\Big|\widehat{A}_{n,\Delta_n}(f_k;\theta)-(n\Delta_n)^{-1}\int_\tau^{n\Delta_n}\mathcal{A}f_k(X_t;\theta)dt\Big|^2 \le L\sum_{j=1}^\infty\Delta_{n_j} < \infty.$$

By Borel–Cantelli lemma, together with the LLN in (17), it follows that

$$\widehat{A}_{n,\Delta_n}(f_k;\theta) \to \int_{\mathscr{C}} \mathcal{A} f_k(\eta;\theta)\mu(d\eta;\theta^*)$$

almost surely for each $\theta \in \Theta$. To prove the uniform convergence, it suffices to show that $\widehat{A}_{n,\Delta}(f_k;\theta)$ is equicontinuous on each sample path. To this end, note that

$$\begin{aligned}
&\sup_{\theta\in\Theta} |\partial_\theta \widehat{A}_{n,\Delta_n}(f_k;\theta)| \le n^{-1}\sum_{i=1}^{n} \sup_{\theta\in\Theta} |\partial_\theta \tilde{b}_{\Delta_n}^{\top}(Z_{i-\ell_{\Delta_n}},\cdots,Z_i;\theta)\nabla f_k(Z_i)| \\
&\le Ln^{-1}\sum_{i=1}^{n}(1+\|X_{i\Delta_n}\|_{\mathscr{C}}^2) \to c
\end{aligned}$$

almost surely for some constant c, which essentially implies that $\widehat{A}_{n,\Delta_n}(f_k;\theta)$ is uniformly Lipschitz on each sample path. The proof is complete.

Proof of Theorem 1 By Proposition 1, it follows that

$$\begin{aligned}
&\sum_{k=1}^{m}\left|\int_{\mathscr{C}} \mathcal{A} f_k(\eta;\hat{\theta}_{n,\Delta_n}(\mathbf{f}))\mu(d\eta;\theta^*)\right| \\
&\le \sum_{k=1}^{m}\sup_{\theta\in\Theta}\left|\widehat{A}_{n,\Delta_n}(f_k;\theta) - \int_{\mathscr{C}} \mathcal{A} f_k(\eta;\theta)\mu(d\eta;\theta^*)\right| + \sum_{k=1}^{m}\left|\widehat{A}_{n,\Delta_n}(f_k;\hat{\theta}_{n,\Delta_n}(\mathbf{f}))\right| \\
&\le \sum_{k=1}^{m}\sup_{\theta\in\Theta}\left|\widehat{A}_{n,\Delta_n}(f_k;\theta) - \int_{\mathscr{C}} \mathcal{A} f_k(\eta;\theta)\mu(d\eta;\theta^*)\right| + \inf_{\theta\in\Theta}\sum_{k=1}^{m}\left|\widehat{A}_{n,\Delta_n}(f_k;\theta)\right| \\
&\le 2\sum_{k=1}^{m}\sup_{\theta\in\Theta}\left|\widehat{A}_{n,\Delta_n}(f_k;\theta) - \int_{\mathscr{C}} \mathcal{A} f_k(\eta;\theta)\mu(d\eta;\theta^*)\right| + \inf_{\theta\in\Theta}\sum_{k=1}^{m}\left|\int_{\mathscr{C}} \mathcal{A} f_k(\eta;\theta)\mu(d\eta;\theta^*)\right| \\
&\to 0.
\end{aligned}$$

Then, the uniqueness of θ^* as the solution to $\int_{\mathscr{C}} \mathcal{A} f_k(\eta;\hat{\theta}_{n,\Delta_n}(\mathbf{f}))\mu(d\eta;\theta^*) = 0$ gives that $\hat{\theta}_{n,\Delta_n}(\mathbf{f}) \to \theta^*$ almost surely.

Proof of Lemma 1 Without loss of generality, it is assumed that

$$\Big(\int_{\mathscr{C}} \partial_\theta \mathcal{A} f_1(\eta;\theta^*)\mu(d\eta;\theta^*),\cdots,\int_{\mathscr{C}} \partial_\theta \mathcal{A} f_k(\eta;\theta^*)\mu(d\eta;\theta^*)\Big) = I,$$

otherwise, a local linear transformation is used. In the small neighborhood of θ^*, the following Taylor's expansion holds

$$\Big(\int_{\mathscr{C}} \mathcal{A} f_1(\eta;\theta)\mu(d\eta;\theta^*),\cdots,\int_{\mathscr{C}} \mathcal{A} f_m(\eta;\theta)\mu(d\eta;\theta^*)\Big) = \theta-\theta^* + o(|\theta-\theta^*|).$$

One can steadily check that the condition for the well-known Poincaré-Miranda theorem is satisfied. As the convergence of $\widehat{A}_{n,\Delta}(f_k;\cdot)$ to $\int_{\mathscr{C}} \mathcal{A} f_k(\eta;\theta)\mu(d\eta;\theta^*)$ is uniform, the Poincaré-Miranda theorem is applicable for $\widehat{A}_{n,\Delta}(f_k;\cdot)$, which yields that $\widehat{A}_{n,\Delta}(f_k;\cdot) = 0$ admits a solution in Θ. The proof is complete.

Proof of Theorem 2 By the definition of $\hat{\theta}_{n,\Delta}(\mathbf{f})$ in (2), note that $\widehat{A}_{n,\Delta}(f_k;\hat{\theta}_{n,\Delta}(\mathbf{f}))=\gamma_{n,\Delta}$. Recall the definition (18) (with $m(\mathcal{A}f_k)=0$) and

$$\sqrt{n\Delta}A_{n\Delta}(f_k;\theta^*)=\frac{1}{\sqrt{n\Delta}}\int_\tau^{n\Delta}\mathcal{A}f_k(X_t;\theta^*)dt,$$

and note that $|\partial_\theta^2\mathcal{A}f_k(\eta;\theta)|\le L(\|\eta\|_{\mathscr{C}}^2+1)$ by Assumption 2. Using Taylor's expansion of θ to obtain the following

$$\begin{aligned}\sqrt{n\Delta}A_{n\Delta}(f_k;\theta^*)&=\sqrt{n\Delta}[A_{n\Delta}(f_k;\theta^*)-A_{n\Delta}(f_k;\hat{\theta}_{n,\Delta}(\mathbf{f})]\\&\quad+\sqrt{n\Delta}[A_{n\Delta}(f_k;\hat{\theta}_{n,\Delta}(\mathbf{f}))-\hat{A}_{n,\Delta}(f_k;\hat{\theta}_{n,\Delta}(\mathbf{f}))]+\sqrt{n\Delta}\gamma_{n,\Delta}\\&=-\sqrt{n\Delta}(\hat{\theta}_{n,\Delta}(\mathbf{f})-\theta^*)^\top[\frac{1}{n\Delta}\int_\tau^{n\Delta}\partial_\theta\mathcal{A}f_k(X_s;\theta^*)ds]\\&\quad+O(1)\Big(\sqrt{n\Delta}|\hat{\theta}_{n,\Delta}(\mathbf{f})-\theta^*|^2\Big[\frac{1}{n\Delta}\int_\tau^{n\Delta}(1+\|X_s\|_{\mathscr{C}}^2)ds\Big]\Big)\\&\quad+O(1)\sqrt{\frac{\Delta}{n}}\sum_{i=\ell_\Delta}^n\Big[w_{\Delta_n}(X_{i\Delta})(1+\|X_{i\Delta}\|_{\mathscr{C}})\Big]+\gamma_{n,\Delta_n}.\end{aligned}\tag{10}$$

In view of (16), then,

$$\begin{aligned}&\sqrt{\frac{\Delta_n}{n}}\sum_{i=\ell_{\Delta_n}}^n\mathbb{E}[w_{\Delta_n}(X_{i\Delta})(1+\|X_{i\Delta_n}\|_{\mathscr{C}})]\\&\le\sqrt{\frac{\Delta_n}{n}}\sum_{i=\ell_{\Delta_n}}^n\sqrt{\mathbb{E}w_\Delta^2(X_{i\Delta_n})(1+\mathbb{E}\|X_{i\Delta_n}\|_{\mathscr{C}}^2)}=O(\sqrt{n}\Delta_n).\end{aligned}$$

Again, an application of (5) yields that

$$(n\Delta_n)^{\frac{1}{2}}|\hat{\theta}_{n,\Delta_n}(\mathbf{f})-\theta^*|^2=o_p(1)(n\Delta_n)^{\frac{1}{2}}(\hat{\theta}_{n,\Delta_n}(\mathbf{f})-\theta^*).$$

The LLN in (17) implies that

$$\frac{1}{n\Delta_n}\int_\tau^{n\Delta_n}(1+\|X_s\|_{\mathscr{C}}^2)ds\to c\text{ and }\frac{1}{n\Delta_n}\int_\tau^{n\Delta_n}\partial_\theta\mathcal{A}f_k(X_s;\theta^*)ds\to r_k$$

almost surely for some constant c. By (10), the asymptotic normality for $A_{n\Delta}(f_k;\theta_0)$ defined in (18) gives that for any $\{\alpha_k:k=1,\cdots,m\}$

$$\sqrt{n\Delta_n}\sum_{k=1}^m\alpha_k\langle r_k(\theta^*),\hat{\theta}_{n,\Delta_n}(\mathbf{f})-\theta^*\rangle\to N\left(0,v^2\left(\sum_{k=1}^m\alpha_k\mathcal{A}_0f_k\right)\right)$$

in distribution. Since $R(\theta^*)=(r_1(\theta^*),\cdots,r_m(\theta^*))$ has a rank of m, there exists a unique $\Sigma(\mathbf{f};\theta^*)$, which is $m\times m$-dimensional non-negative definite, symmetric matrix such that $r_k^\top\Sigma(\mathbf{f};\theta^*)r_k=v^2(\mathcal{A}_0f_k;\theta^*)$ for all $1\le k\le m$ given in (7). Then, (6) holds for such $\Sigma(\mathbf{f};\theta^*)$. The proof is established.

Proof of Proposition 2 By Lemma 2 and the consistency of $\hat{\theta}_{n,\Delta_n}(\mathbf{f})$, it follows that $\Sigma^{-1/2}(\mathbf{f};\hat{\theta}_{n,\Delta_n}(\mathbf{f}))\to\Sigma^{-1/2}(\mathbf{f};\theta^*)$ in probability. Therefore, Proposition 2 is a direct consequence of Theorem 2.

Proof of Theorem 3 Note that

$$\hat{A}_{n,\Delta}(f_0;\hat{\theta}_{n,\Delta}(\mathbf{f})) = \Big(\hat{A}_{n,\Delta}(f_0;\hat{\theta}_{n,\Delta}(\mathbf{f})) - \hat{A}_{n,\Delta}(f_0;\theta^*)\Big) + \Big(\hat{A}_{n,\Delta}(f_0;\theta^*) - A_{n\Delta}(f_0;\theta^*)\Big) + A_{n\Delta}(f_0;\theta^*). \tag{11}$$

The Itô formula yields that

$$A_{n\Delta}(f_0;\theta^*) = \frac{1}{\sqrt{n\Delta}}\Big(f_0(X(t)) - f_0(X(\tau)) - \int_\tau^{n\Delta}\sigma^\top(X_t)\nabla f_0(X(t))dW_t\Big). \tag{12}$$

Also, note that

$$\begin{aligned}
&\mathbb{E}\big|\hat{A}_{n,\Delta}(f_0;\theta^*) - A_{n\Delta}(f_0;\theta^*)\big| \\
&\leq \sqrt{\frac{\Delta}{n}}\mathbb{E}\Big|\sum_{i=\ell_\Delta}^n\Big[\tilde{b}_\Delta^\top(Z_{i-\ell_\Delta},\cdots,Z_i;\hat{\theta}_{n,\Delta}(\mathbf{f}))\nabla f_0(Z_i) \\
&\quad + \frac{1}{2}\mathrm{trace}\Big([\tilde{\sigma}_\Delta^\top\sigma_\Delta](Z_{i-\ell_\Delta},\cdots,Z_i)D^2 f_0(Z_i)\Big)\Big] \\
&\quad - \sum_{i=\ell_\Delta}^n\Big[b^\top(X_{i\Delta};\theta^*)\nabla f_0(X(i\Delta)) + \frac{1}{2}\mathrm{trace}\Big([\sigma^\top\sigma]X_{i\Delta})D^2 f_0(X(i\Delta))\Big)\Big]\Big| \\
&\quad + \frac{1}{\sqrt{n\Delta}}\mathbb{E}\left(\sum_{i=\ell_\Delta}^n\Big|\int_{i\Delta}^{(i+1)\Delta} b^\top(X_t;\theta^*)\nabla f_0(X(t)) - b^\top(X_{i\Delta};\theta^*)\nabla f_0(X(i\Delta))dt\Big|\right) \\
&\quad + \frac{1}{2\sqrt{n\Delta}}\mathbb{E}\left(\sum_{i=\ell_\Delta}^n\Big|\mathrm{trace}\Big(\int_{i\Delta}^{(i+1)\Delta}[\sigma^\top\sigma](X_t)D^2 f_0(X(t))\right. \\
&\quad \left. - [\sigma^\top\sigma](X_{i\Delta})D^2 f_0(X(i\Delta))\Big)dt\Big|\right) \\
&\leq L\sqrt{\frac{\Delta}{n}}\mathbb{E}\left(\sum_{i=\ell_\Delta}^n w_\Delta(X_{i\Delta})\cdot(\|X_{i\Delta}\|_{\mathscr{C}}+1)\right) \\
&\leq L\sqrt{\frac{\Delta}{n}}\sum_{i=\ell_\Delta}^n\sqrt{\mathbb{E}w_\Delta^2(X_{i\Delta})\cdot(1+\mathbb{E}\|X_{i\Delta}\|_{\mathscr{C}}^2)} \leq L\sqrt{n}\Delta.
\end{aligned} \tag{13}$$

Using Taylor's expansion, one has

$$\begin{aligned}
&\hat{A}_{n,\Delta}(f_0;\hat{\theta}_{n,\Delta}(\mathbf{f})) - \hat{A}_{n,\Delta}(f_0;\theta^*) \\
&= \sqrt{\frac{\Delta}{n}}\sum_{i=\ell_\Delta}^n\Big([\tilde{b}_\Delta(Z_{i-\ell_\Delta},\cdots,Z_i;\hat{\theta}_{n,\Delta}(\mathbf{f})) - \tilde{b}_\Delta(Z_{i-\ell_\Delta},\cdots,Z_i;\theta^*)]^\top\nabla f_0(Z_i)\Big) \\
&= \sqrt{n\Delta}\cdot M_{n,\Delta}^\top(\hat{\theta}_{n,\Delta}(\mathbf{f}) - \theta^*) + O(1)\sqrt{n\Delta}|\hat{\theta}_{n,\Delta}(\mathbf{f}) - \theta^*|^2\cdot\Big[n^{-1}\sum_{i=\ell_\Delta}^n(1+\|X_{i\Delta}\|_{\mathscr{C}})\Big] \\
&= M_{n,\Delta}^\top[R^{-1}(\theta^*)]^\top\Big[\sqrt{n\Delta}R^\top(\theta^*)(\hat{\theta}_{n,\Delta}(\mathbf{f}) - \theta^*)\Big] \\
&\quad + O(1)\sqrt{n\Delta}|\hat{\theta}_{n,\Delta}(\mathbf{f}) - \theta^*|^2\cdot\Big[n^{-1}\sum_{i=\ell_\Delta}^n(1+\|X_{i\Delta}\|_{\mathscr{C}})\Big]
\end{aligned}$$

$$= M_{n,\Delta}^{\top}[R^{-1}(\theta^*)]^{\top}\left[\frac{1}{\sqrt{n\Delta}}\int_{\tau}^{n\Delta}\begin{pmatrix}\sigma^{\top}(X_t)\nabla f_1(X(t))\\ \vdots \\ \sigma^{\top}(X_t)\nabla f_m(X(t))\end{pmatrix}dW_t\right] + o_p(1)$$

$$+ O(1)\sqrt{n\Delta}|\hat{\theta}_{n,\Delta}(\mathbf{f}) - \theta^*|^2 \cdot \Big[n^{-1}\sum_{i=\ell_\Delta}^{n}(1 + \|X_{i\Delta}\|_{\mathscr{C}})\Big], \tag{14}$$

where

$$M_{n,\Delta} = \frac{1}{n}\left\{\sum_{i=\ell_\Delta}^{n}\partial_\theta \tilde{b}_\Delta^{\top}(Z_{i-\ell_\Delta},\cdots,Z_i;\theta^*)\nabla f_0(Z_i)\Big]\right\}.$$

By the LLN, $M_{n,\Delta_n} \to \int_{\mathscr{C}} \partial_\theta \mathcal{A} f_0(\eta;\theta^*)\mu(d\eta;\theta^*)$ almost surely as $n \to \infty$. Lemma 2 together with (11)–(14) implies that

$$\hat{A}_{n,\Delta_n}(f_0;\hat{\theta}_{n,\Delta_n}(\mathbf{f})) = \langle R^{-1}(\theta^*)r_0(\theta^*), \frac{1}{\sqrt{n\Delta_n}}\int_{\tau}^{n\Delta}\begin{pmatrix}\sigma^{\top}(X_t)\nabla f_1(X(t))\\ \vdots \\ \sigma^{\top}(X_t)\nabla f_m(X(t))\end{pmatrix}dW_t\rangle$$

$$- \frac{1}{\sqrt{n\Delta_n}}\int_{\tau}^{n\Delta_n}\sigma^{\top}(X_t)\nabla f_0(X(t))dW_t + o_p(1).$$

Therefore, our central limit theorem holds.

Proof of Theorem 4 For this case $\hat{\theta}_{n,\Delta} = \theta^* = \theta_0$, the asymptotic normality in (9) follows the same way as the proof of Theorem 3. Under H_1, the testing statistic satisfies

$$\frac{1}{\sqrt{n\Delta_n}}A_{n,\Delta_n}(f_0,\theta^*) \to \int_{\mathscr{C}} \mathcal{A} f_0(\eta;\theta^*)\mu(d\eta;\theta_1)$$

almost surely by the LLN. Moreover, using the exponential ergodicity of SDDE in (1) for $\theta = \theta_1$, one obtains

$$\lim_{n,\Delta}\mathbb{E}\Big[A_{n,\Delta_n}(f_0,\theta^*) - \sqrt{n\Delta_n}\int_{\mathscr{C}}\mathcal{A} f_0(\eta(0),\eta;\theta^*)\mu(d\eta;\theta_1)\Big]^2 < \infty.$$

Then, Chebyshev's inequality yields that

$$\begin{aligned}&\mathbb{P}\Big(|v^{-1}(\mathcal{A} f_0;\theta^*)\cdot A_{n,\Delta_n}(f_0,\theta^*)| \le z_{\alpha/2}\Big)\\ &\le \mathbb{P}\Big(\Big|v^{-1}(\mathcal{A} f_0;\theta^*)\cdot[A_{n,\Delta_n}(f_0,\theta^*) - \sqrt{n\Delta_n}\int_{\mathscr{C}}\mathcal{A} f_0(\eta(0),\eta;\theta^*)\mu(d\eta;\theta_1)]\Big|\\ &\ge L\sqrt{n\Delta_n} - z_{\alpha/2}\Big) \le L(n\Delta_n)^{-1},\end{aligned}$$

which implies that the probability of Type II error converges to 0 with a rate of $(n\Delta_n)^{-1}$. In other words, the test power converges to one. Therefore, the proof is complete.

Declarations

Statement on Conflict of interest The authors claim that there is no conflict of interest about the manuscript submitted.

Appendix: General results on SDDEs

In this appendix, recall the ergodicity theory for SDDEs for our problem from Bao et al. (2020). Note that the following theorem is from Bao et al. (2020) concerning about the exponential ergodicity of SDDEs.

Theorem 5 *Suppose Assumption 1 holds. Then, the followings are true.*
(i) *The Markov process $\{X_t\}$ admits a unique invariant measure μ on $\mathscr{C}$ with for any $p \geq 1$*

$$\sup_{t\geq 0} \mathbb{E}\|X_t\|_{\mathscr{C}}^{2p} < L_p, \tag{15}$$

and

$$\sup_{t\geq 0} \delta^{-p} \mathbb{E} w_\delta^{2p}(X_t) < L_p, \tag{16}$$

where L_p is a constant independent of δ.
(ii) *If $|g(\eta)| \leq L\|\eta\|_{\mathscr{C}}^2$ for some $L > 0$, the following law of large numbers holds*

$$\frac{1}{T}\int_\tau^T g(X_t)dt \to m(g) = \int_{\mathscr{C}} g(\eta)\mu(d\eta;\theta) \tag{17}$$

almost surely.
(iii) *For any $h : \mathscr{C} \mapsto \mathbb{R}$ satisfying*

$$|h(\eta) - h(\xi)| \leq L\|\eta - \xi\|_{\mathscr{C}},$$

one has

$$A_T(h;\theta^*) = \frac{1}{\sqrt{T}}\int_\tau^T [h(X_t) - m(h)]dt \to N\left(0, v^2(h;\theta)\right) \tag{18}$$

in distribution, where X_t^η is the solution to (1) with initial $X_0 = \eta$,

$$R_f(\eta) = \int_0^\infty \mathbb{E}f(X_t^\eta) - m(f)dt,$$

and

$$v^2(h;\theta) = \int_{\mathscr{C}} \mu(d\eta;\theta)\left[\mathbb{E}\left|\int_0^1 f(X_t^\eta)dt + R_f(X_1^\eta) - R_f(\eta)\right|\right]^2. \tag{19}$$

In particular, if $h(\eta) = \mathcal{A}f(\eta(0), \eta; \theta^*)$ some twice continuously differentiable f with bounded second order derivatives, one has

$$v^2(\mathcal{A}_0 f;\theta^*) = \int_{\mathscr{C}} |\sigma^\top(\eta;\theta^*)\nabla f(\eta(0))|^2 \mu(d\eta;\theta^*). \tag{20}$$

Remark 2 Our statements (i) and (ii) are from (1.2) and Statement (2) in Theorem 1.1 from Bao et al. (2020), respectively. The statement (iii) is from Statement (1) in Theorem 1.2 from Bao et al. (2020).

For our testing problem, we finish the appendix with the following lemma concerning with the continuity of the invariant measure $\mu(\cdot;\theta)$ with respect to θ.

Lemma 2 *Assume Assumption 1 holds and $\sup_{\theta\in\Theta}\sup_{t\geq 0}\mathbb{E}\|X_t\|_{\mathscr{C}}^2 < \infty$. Further, assume that $|\sigma(\xi,\theta) - \sigma(\eta,\theta)| \leq \lambda_3|\xi - \eta|_{\mathscr{C}}$ with $\lambda_1 > (\lambda_2 + \lambda_3^2)e^{-\lambda_1\tau}$. Then, as $\theta \to \theta^*$, $\mu(\cdot;\theta) \to \mu(\cdot;\theta^*)$ in distribution with*

$$\int_{\mathscr{C}} \|\eta\|_{\mathscr{C}}^2 \mu(d\eta;\theta) \to \int_{\mathscr{C}} \|\eta\|_{\mathscr{C}}^2 \mu(d\eta;\theta^*). \tag{21}$$

Proof Suppose $X(t)$ and $Y(t)$ be the solution of SDDE (1) with same initial and $\theta = \theta^*$ and θ_1 respectively. Note that

$$\begin{aligned} d(X(t) - Y(t)) &= [b(X_t;\theta^*) - b(Y_t;\theta_1)]dt + [\sigma(X_t;\theta^*) - \sigma(Y_t;\theta_1)]dW(t) \\ &= [b(X_t;\theta^*) - b(Y_t;\theta^*)]dt + [\sigma(X_t;\theta^*) - \sigma(Y_t;\theta^*)]dW(t) \\ &\quad +[b(Y_t;\theta^*) - b(Y_t;\theta_1)]dt + [\sigma(Y_t;\theta^*) - \sigma(Y_t;\theta_1)]dW(t). \end{aligned}$$

Therefore, taking $\delta > 0$ such that $\lambda_1 > (\lambda_2 + \lambda_3^2)e^{-\lambda_1\tau}$ yields that

$$\begin{aligned} d|X(t) - Y(t)|^2 &= \Big[2\langle X(t) - Y(t), b(X_t;\theta^*) - b(Y_t;\theta^*)\rangle + |\sigma(X_t;\theta^*) - \sigma(Y_t;\theta^*)|^2 \\ &\quad +\Big[2\langle X(t) - Y(t), b(Y_t;\theta^*) - b(Y_t;\theta_1)\rangle + |\sigma(Y_t;\theta^*) - \sigma(Y_t;\theta_1)|^2 \\ &\quad +2(\sigma(Y_t;\theta^*) - \sigma(Y_t;\theta_1))(\sigma(X_t;\theta^*) - \sigma(Y_t;\theta^*))dt + dM \\ &\le -\lambda_1|X(t) - Y(t)|^2 + \lambda_2\|X_t - Y_t\|_{\mathscr{C}}^2 + \lambda_3^2\|X_t - Y_t\|_{\mathscr{C}}^2 \\ &\quad +2L\|X_t - Y_t\|_{\mathscr{C}}|\theta^* - \theta_1|\|Y_t\|_{\mathscr{C}} + L|\theta^* - \theta_1|^2 + L|\theta^* - \theta_1|\|X_t - Y_t\|_{\mathscr{C}} + dM \\ &\le -\lambda_1|X(t) - Y(t)|^2 + (\lambda_2 + \lambda_3^2)\|X_t - Y_t\|_{\mathscr{C}}^2 + L|\theta^* - \theta_1|(1 + \|X_t\|_{\mathscr{C}}^2 + \|Y_t\|_{\mathscr{C}}^2) + dM, \end{aligned}$$

where M is a martingale. Similar to the proof of Lemma 3.1 in Bao et al. (2020), one has

$$\lim_{t\to\infty} \mathbb{E}\|X_t - Y_t\|_{\mathscr{C}}^2 \le L_\delta|\theta_1 - \theta^*|.$$

As (X_t, Y_t) is an asymptotic coupling of $\mu(\cdot;\theta^*)$ and $\mu(\cdot;\theta_1)$, this proves that $\mu(\cdot;\theta_1) \to \mu(\cdot;\theta^*)$ in distribution and (21) holds. The proof is complete. □

Finally, a remark should be mentioned here that the condition in the above lemma is sufficient but not necessary. How to get a better condition is beyond the scope of our paper, thus omitted here, and it deserves a further investigation.

References

Aït-Sahalia Y (1996) Testing continuous-time models of the spot interest rate. Rev Financ Stud 9(2):385–426

Aït-Sahalia Y, Fan J, Peng H (2009) Nonparametric transition-based tests for jump diffusions. J Am Stat Assoc 104(487):1102–1116

Arriojas M, Hu Y, Mohammed SE, Pap G (2007) A delayed Black and Scholes formula. Stoch Anal Appl 25(2):471–492

Bao J, Wang FY, Yuan C (2020) Limit theorems for additive functionals of path-dependent SDEs. Discrete Contin Dyn Syst A 40(9):5173–5188

Benke JM, Pap G (2017) One-parameter statistical model for linear stochastic differential equation with time delay. Statistics 51(3):510–531

Bratsun D, Volfson D, Tsimring LS, Hasty J (2005) Delay-induced stochastic oscillations in gene regulation. Proc Natl Acad Sci 102(41):14593–14598

Chen L, Yu Z (2014) Maximum principle for nonzero-sum stochastic differential game with delays. IEEE Trans Autom Control 60(5):1422–1426

Chen SX, Gao J, Tang CY (2008) A test for model specification of diffusion processes. Ann Stat 36(1):167–198

Dachian S, Kutoyants YA (2008) On the goodness-of-fit tests for some continuous time processes. In: Vonta F et al (eds) Statistical models and methods for biomedical and technical systems. Birkhäuser, Boston, pp 385–403

Gushchin AA, Küchler U (1999) Asymptotic inference for a linear stochastic differential equation with time delay. Bernoulli 5(6):1059–1098

Hobson DG, Rogers LC (1998) Complete models with stochastic volatility. Math Financ 8(1):27–48

Hong Y, Li H (2005) Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Rev Financ Stud 18(1):37–84

Ivanov AF, Swishchuk AV (2008) Optimal control of stochastic differential delay equations with application in economics. Int J Qualit Theory Differ Equ Appl 2(2):201–213
Karatzas I (1996) Anticipative portfolio optimization. Adv Appl Probab 28(4):1095–1122
Kleptsyna M, Kutoyants YA (2014) On asymptotically distribution free tests with parametric hypothesis for ergodic diffusion processes. Stat Infer Stoch Process 17(1):295–319
Küchler U, Sørensen M (2010) A simple estimator for discrete-time samples from affine stochastic delay differential equations. Stat Infer Stoch Process 13(1):125–132
Küchler U, Kutoyants YA (2000) Delay estimation for some stationary diffusion-type processes. Scand J Stat 27(3):405–414
Küchler U, Sørensen M (2013) Statistical inference for discrete-time samples from affine stochastic delay differential equations. Bernoulli 19(2):409–425
Kutoyants YA (2021) On multi-step estimation of delay for SDE. Bernoulli 27(3):2069–2090
Lawrence DB (2012) The economic value of information. Springer, New York
Lei J, Mackey MC (2007) Stochastic differential delay equation, moment stability, and application to hematopoietic stem cell regulation system. SIAM J Appl Math 67(2):387–407
López-Prez A, Febrero-Bande M, Gonzàlez-Manteiga W (2022) A comparative review of specification tests for diffusion models. arXiv:2208.08420
Marschak J (1971) Economics of information systems. J Am Stat Assoc 66(333):192–219
Mao X (2007) Stochastic differential equations and applications. Elsevier, Amsterdam
Negri I, Nishiyama Y (2009) Goodness of fit test for ergodic diffusion processes. Ann Inst Stat Math 61(4):919–928
Reiss M (2005) Adaptive estimation for affine stochastic delay differential equations. Bernoulli 11(1):67–102
Rihan FA (2021) Delay differential equations and applications to biology. Springer, Singapore
Steiner J, Stewart C, Matějka F (2017) Rational inattention dynamics: Inertia and delay in decision-making. Econometrica 85(2):521–553
Stoica G (2005) A stochastic delay financial model. Proc Am Math Soc 133(6):1837–1841
Vasicek O (1977) An equilibrium characterization of the term structure. J Financ Econ 5(2):177–188