Electronic Journal of Statistics
Vol. 14 (2020) 3914–3938
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1761

Weak convergence of marked empirical processes in a Hilbert space and its applications

Koji Tsukuda[1] **and Yoichi Nishiyama**[2]

[1] *Faculty of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka-shi, Fukuoka 819-0395, Japan. e-mail:* tsukuda@math.kyushu-u.ac.jp

[2] *Faculty of International Research and Education, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, Japan. e-mail:* nishiyama@waseda.jp

Abstract: We discuss goodness-of-fit tests for stationary ergodic processes based on limit theorems for marked empirical processes viewed as elements of an L^2 space. Our limit theorems cover a weighted process that enables an Anderson–Darling-type test statistic for the goodness-of-fit tests to be proposed. The procedures presented for these goodness-of-fit tests are novel.

MSC2020 subject classifications: Primary 62M02; secondary 60F05, 60F17, 60G42, 62G10.
Keywords and phrases: Diffusion process, nonlinear time series, goodness-of-fit test, weak convergence in Hilbert space.

Received February 2020.

1. Introduction and limit theorems

With regard to the goodness-of-fit tests for stationary ergodic processes, we propose a Cramér–von Mises type statistic based on discrete time observations to test a simple hypothesis for a diffusion process and an Anderson–Darling-type statistic for a nonlinear time series. For that purpose, we provide two limit theorems, Theorems 1.1 and 1.2 presented in Sections 1.1 and 1.2, which assert weak convergence of marked empirical processes in L^2 space.

Goodness-of-fit tests have been extensively studied in the literature because they are useful in deciding whether a mathematical model is acceptable to describe sampled data. We refer to the work of González-Manteiga and Crujeiras [5] for a review on goodness-of-fit tests, in which Section 5 is devoted to tests when dependent sequences are observed. Among the abundant work treating goodness-of-fit tests for stochastic process models, we are interested in an approach based on empirical processes marked by residuals. This approach has been developed by Koul and Stute [10] and Escanciano [3] among others. The main objective of this paper is proposing new test procedures that may not be justified by limit theorems stated in [10] and [3]. Specifically, our limit theorems do *not* include Theorem 2.1 of [10] or Theorem 1 of [3], but our results contain the following merits. The assumptions supposed in either of these theorems do not suit the setting of our diffusion process; nevertheless, Theorem 1.1 does applied. Moreover, although a weak convergence of an Anderson–Darling-type test statistic cannot be directly derived from the weak convergence in the Skorokhod space or ℓ^∞ space, Theorem 1.2 enables us to consider the Anderson–Darling-type test statistic. Indeed, the applications described in this paper are novel.

Remark 1.1. Some goodness-of-fit tests for ergodic diffusion processes have been extensively studied. The first study on this topic is apparently Kutoyants [11] (Section 5.4). This and subsequent work such as Dachian and Kutoyants [1], Kleptsyna and Kutoyants [9], Kutoyants [12, 13], Negri and Nishiyama [15] treat the problem with continuous time observations. However, from the viewpoint of applications, discrete time observations need to be considered. Using a smoothed empirical process, Masuda et al. [14] proposed Kolmogorov–Smirnov-type goodness-of-fit tests based on discrete time observations. They established not the weak convergence of $Z_n(x)$, which shall be defined in (1.1), but the weak convergence of its smoothed version. See also Negri and Nishiyama [16] for a review on goodness-of-fit tests for ergodic diffusion processes.

Remark 1.2. As for goodness-of-fit test for time series models, Koul and Stute [10] considered not only a simple hypothesis but also one that is a parametric composite based on the notion of the martingale transformation (Khmaladze [8]). Moreover, as for ergodic diffusion process models (continuous time observations), Kleptsyna and Kutoyants [9] and Kutoyants [12, 13] considered a parametric composite hypothesis, although marked empirical processes were not treated. Nevertheless, we only consider a simple hypotheses; composite hypotheses are set aside as a possible direction of study in future research.

Next, we describe our limit theorems. Let $L^2(\mathbb{R},\nu)$ be the set of equivalence classes of square integrable functions on $\mathbb{R}$ with respect to a finite Borel measure ν on $\mathbb{R}$. The $L^2(\mathbb{R},\nu)$ space is equipped with an inner product $\langle\cdot,\cdot\rangle$ defined by $\langle f,g\rangle = \int_{\mathbb{R}} f(x)g(x)\nu(dx)$ for $f,g \in L^2(\mathbb{R},\nu)$ and norm $\|\cdot\|$ defined by $\|f\| = \langle f,f\rangle^{1/2}$ for $f \in L^2(\mathbb{R},\nu)$. For any interval A, the function $1_A(\cdot)$ is defined by $1_A(x) = 1$ $(x \in A)$, 0 $(x \notin A)$. For every positive integer n, we introduce a filtered probability space $(\Omega^n, \mathcal{F}^n, \mathbf{F}^n = \{\mathcal{F}_i^n\}_{i\geq 0}, P^n)$. Let $\{X_i^n\}_{i\geq 0}$ be a real-valued $\mathbf{F}^n$-adapted sequence and $\{m_i^n\}_{i\geq 1}$ a real valued $\mathbf{F}^n$-adapted martingale difference sequence (thus for all i, m_i^n is $\mathcal{F}_i^n$-measurable and $E^n[m_i^n|\mathcal{F}_{i-1}^n] = 0$ almost surely). We study the asymptotic behavior of an empirical process marked by the martingale difference sequence $\{m_i^n\}_{i\geq 1}$

$$x \rightsquigarrow Z_n(x) = \sum_{i=1}^n 1_{(-\infty,x]}(X_{i-1}^n)m_i^n \tag{1.1}$$

and its weighted process with weight function $x \mapsto w(x)(>0)$

$$x \rightsquigarrow Z_n^w(x) = w(x)Z_n(x) = \sum_{i=1}^n w(x)1_{(-\infty,x]}(X_{i-1}^n)m_i^n.$$

Their limits are Gaussian processes G and G^w defined by

$$x \rightsquigarrow G(x) = B(\Psi(x)),$$

and

$$x \rightsquigarrow G^w(x) = w(x)B(\Psi(x)),$$

respectively, where $x \rightsquigarrow B(x)$ is a standard Brownian motion and Ψ is the limit (for the exact sense of the limit; see Assumptions 1.1 and 1.2) of

$$x \rightsquigarrow \sum_{i=1}^{n} 1_{(-\infty,x]}(X_{i-1}^n)E^n[(m_i^n)^2|\mathcal{F}_{i-1}^n].$$

To explain the problem simply, let us consider the instance when $(X_{i-1}^n)_{i\geq 1} = (X_{i-1})_{i\geq 1}$ and $(m_i^n)_{i\geq 1} = (m_i/\sqrt{n})_{i\geq 1}$ are independent sequences of independent and identically distributed random variables satisfying the following:

- X_0 is an absolutely continuous random variable for which the distribution function is F_X,
- $E[m_1] = 0$, $E[(m_1)^2] < \infty$.

If

$$\int_{\mathbb{R}} w(x)^2 F_X(x)\nu(dx) < \infty, \tag{1.2}$$

then the central limit theorem in a separable Hilbert space (see Example 1.8.5 of van der Vaart and Wellner [19]) implies that

$$x \rightsquigarrow Z_n^w(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w(x)1_{(-\infty,x]}(X_{i-1})m_i$$

converges weakly to

$$x \rightsquigarrow G^w(x) = \sqrt{E[(m_1)^2]}w(x)B(F_X(x))$$

in $L^2(\mathbb{R},\nu)$ as $n \to \infty$, because

$$E[\|w1_{(-\infty,\cdot]}(X_0)m_1\|^2] = E[(m_1)^2]\int_{\mathbb{R}} w(x)^2 F_X(x)\nu(dx) < \infty$$

and

$$E\left[\langle w1_{(-\infty,\cdot]}(X_0)m_1, h\rangle^2\right] = E\left[\left\langle \sqrt{E[(m_1)^2]}wB(F_X), h\right\rangle^2\right]$$

for any $h \in L^2(\mathbb{R},\nu)$. Here, the weak convergence in $L^2(\mathbb{R},\nu)$ is defined by

$$E[f(Z_n^w)] \to E[f(G^w)]$$

for any continuous and bounded function $f : L^2(\mathbb{R},\nu) \to \mathbb{R}$. From this weak convergence and the continuous mapping theorem, we conclude that $\|Z_n^w\|^2$ converges in distribution to $\|G^w\|^2$ as $n \to \infty$. Note that when $w(x) = 1$ ($x \in \mathbb{R}$), (1.2) holds automatically. Hence, our contribution is generalizing the simple conditions above to more complicated ones that are more useful when conducting statistical inference for stochastic processes. The limit theorems are presented in the following two subsections.

1.1. Weak convergence of Z_n

We first provide the limit theorem for Z_n. The following assumption is a sufficient condition to show the weak convergence of Z_n.

Assumption 1.1. *(i) As $n \to \infty$,*

$$\sum_{i=1}^{n} 1_{(-\infty,x]}(X_{i-1}^n) E^n[(m_i^n)^2 | \mathcal{F}_{i-1}^n] \to^p \Psi(x) \quad (\forall x \in \mathbb{R}) \tag{1.3}$$

holds, where $x \mapsto \Psi(x)$ is a continuous nondecreasing function on $\mathbb{R}$ satisfying $\Psi(x) \downarrow 0$ as $x \to -\infty$ and $\Psi(x) \uparrow \Psi(\infty) < \infty$ as $x \to \infty$, and $\to^p$ denotes convergence in probability.

(ii) There exists a constant $\delta > 0$ such that

$$\sum_{i=1}^{n} E^n[|m_i^n|^{2+\delta} | \mathcal{F}_{i-1}^n] \to^p 0$$

as $n \to \infty$.

(iii) There exists a measurable function ϕ on $\mathbb{R}$ such that for all $n \in \mathbb{N}$ there exist some nonnegative constants c_i^n $(i = 1, \dots, n)$ satisfying

$$\sup_n \left(\sum_{i=1}^{n} c_i^n \right) < \infty$$

and $E^n[(m_i^n)^2 | \mathcal{F}_{i-1}^n] \le c_i^n \phi(X_{i-1}^n)$ almost surely.

(iv) All X_i^n's have the same distribution as ζ such that $E[\phi(\zeta)] < \infty$.

The following limit theorem asserts the weak convergence of Z_n, the proof of which is presented in Appendix A.

Theorem 1.1. *Under Assumption 1.1, Z_n converges weakly to G in $L^2(\mathbb{R}, \nu)$ as $n \to \infty$.*

Remark 1.3. An important point of Theorem 1.1 is that we avoid the assumption (B) in Lemma 3.1 of Koul and Stute [10], which places a restriction on the transition density of a discrete time Markovian process and hence is not suitable for our diffusion process model considered in Section 2. Although Escanciano [3] gave a result for a non-Markovian process, he assumed a condition on the smoothness (the condition (D) in his Theorem 1) of the model that also is not suitable for our purpose. However, notice that our result does not cover theirs because they considered the weak convergence under a uniform metric.

1.2. Weak convergence of Z_n^w

We next provide a limit theorem for Z_n^w. The following assumption is a sufficient condition to show the weak convergence of Z_n^w. Obviously, if we set $w(\cdot) = 1$, then Z_n^w becomes Z_n. However, Theorem 1.1 has been separately stated, because (1.4) is stronger than (1.3).

Assumption 1.2. *(i) As $n \to \infty$,*

$$E^n\left[\left|\sum_{i=1}^n 1_{(-\infty,x]}(X_{i-1}^n)E^n[(m_i^n)^2|\mathcal{F}_{i-1}^n] - \Psi(x)\right|\right] \to 0 \quad (\forall x \in \mathbb{R}) \tag{1.4}$$

holds, where $x \mapsto \Psi(x)$ is a continuous nondecreasing function on $\mathbb{R}$ satisfying $\Psi(x) \downarrow 0$ as $x \to -\infty$, $\Psi(x) \uparrow \Psi(\infty) < \infty$ as $x \to \infty$. Moreover, there exists a nondecreasing function Φ such that

$$I_n(x) \leq \Phi(x) \quad (\forall x \in \mathbb{R}) \tag{1.5}$$

for all sufficiently large n, where

$$I_n(x) = \sum_{i=1}^n E^n\left[1_{(-\infty,x]}(X_{i-1}^n)(m_i^n)^2\right].$$

Furthermore, the following condition holds

$$\int_{\mathbb{R}} (\Psi(x) + \Phi(x))(w(x))^2 \nu(dx) < \infty.$$

(ii) There exists a constant $\delta > 0$ such that

$$\sum_{i=1}^n E^n[1_{(-\infty,x]}(X_{i-1}^n)|m_i^n|^{2+\delta}] \to 0 \quad (\forall x \in \mathbb{R})$$

as $n \to \infty$, and there exists a function Λ such that

$$\sum_{i=1}^n E^n[1_{(-\infty,x]}(X_{i-1}^n)|m_i^n|^{2+\delta}] \leq \Lambda(x) \quad (\forall x \in \mathbb{R}) \tag{1.6}$$

for all sufficiently large n and that

$$\int_{\mathbb{R}} \Lambda(x)(w(x))^{2+\delta} \nu(dx) < \infty.$$

The following limit theorem asserts the weak convergence of Z_n^w. Its proof is presented in Appendix B. From a practical viewpoint, the case where $w(\cdot) = (\Psi(\cdot))^{-1/2}$ is important because it corresponds to a standardization.

Theorem 1.2. *Under Assumption 1.2, Z_n^w converges weakly to G^w in $L^2(\mathbb{R}, \nu)$ as $n \to \infty$.*

Remark 1.4. As for (1.4), it follows from a well-known fact on the uniform integrability that if

$$\sum_{i=1}^n 1_{(-\infty,x]}(X_{i-1}^n)E^n[(m_i^n)^2|\mathcal{F}_{i-1}^n] \to \Psi(x) \quad \text{a.s.} \quad (\forall x \in \mathbb{R}),$$

then (1.4) is equivalent to the uniform integrability of

$$\sum_{i=1}^{n} 1_{(-\infty,x]}(X_{i-1}^n) E^n[(m_i^n)^2 | \mathcal{F}_{i-1}^n]$$

for all $x \in \mathbb{R}$, and also equivalent to $I_n(x) \to \Psi(x)$ for all $x \in \mathbb{R}$.

Remark 1.5. As for (1.5), if we assume Assumptions 1.1 (iii) and (iv), then

$$\begin{aligned} I_n(x) &= E^n\left[\sum_{i=1}^{n} 1_{(-\infty,x]}(X_{i-1}^n) E^n[(m_i^n)^2 | \mathcal{F}_{i-1}^n]\right] \\ &\leq E^n\left[\sum_{i=1}^{n} 1_{(-\infty,x]}(X_{i-1}^n) c_i^n \phi(X_{i-1}^n)\right] \\ &= E\left[\sum_{i=1}^{n} 1_{(-\infty,x]}(\zeta) c_i^n \phi(\zeta)\right] \\ &\leq \left(\sup_n \sum_{i=1}^{n} c_i^n\right) E\left[1_{(-\infty,x]}(\zeta)\phi(\zeta)\right], \end{aligned}$$

and hence we can take $\Phi(x)$ as the right-hand side of the above display (if the integrability condition holds).

2. Cramér–von Mises type goodness-of-fit test for drift parameters in diffusion processes

We now consider the goodness-of-fit test procedure for a diffusion process model. The procedure is justified using Theorem 1.1.

2.1. Problem setting and test procedure

We consider a strictly stationary ergodic stochastic process $\{X_t\}_{t\geq 0}$ which is a solution to a one-dimensional stochastic differential equation (SDE)

$$X_t = X_0 + \int_0^t S(X_s)ds + \int_0^t \sigma(X_s)dW_s \quad (t \geq 0), \tag{2.1}$$

where the random variable X_0 is an almost surely finite initial value, $S(\cdot)$ a measurable function of interest, $\sigma(\cdot)$ a known positive measurable function, and $t \rightsquigarrow W_t$ a standard Wiener process defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, P)$. Let us list the assumptions on S and σ.

(A1) There exists a constant $C > 0$ such that

$$|S(x) - S(y)| \leq C|x - y|, \quad |\sigma(x) - \sigma(y)| \leq C|x - y| \quad (\forall x, y \in \mathbb{R}).$$

(A2) The process $(X_t)_{t\in[0,\infty)}$ is a solution to the SDE (2.1) for (S, σ) and it is stationary and ergodic with absolutely continuous invariant law $\mu_{S,\sigma}$ (that is, $t^{-1}\int_0^t g(X_s)ds \to^p \int_{\mathbb{R}} g(x)\mu_{S,\sigma}(dx)$ as $t\to\infty$ for every $\mu_{S,\sigma}$-integrable function g). We also assume that

$$\int_{\mathbb{R}} |x|^3 \mu_{S,\sigma}(dx) < \infty.$$

Remark 2.1. Assumption **(A1)** implies that there exists a constant $C' > 0$ such that $|S(x)| \le C'(1+|x|)$ and $|\sigma(x)| \le C'(1+|x|)$.

In our problem, from the continuous stochastic process (2.1), $\{X_{t_i^n}\}_{i=1}^n$ is observed at discrete time points $0 = t_0^n < t_1^n < \cdots < t_n^n$ satisfying

$$t_n^n \to \infty, \quad n\Delta_n^2 \to 0 \tag{2.2}$$

as $n \to \infty$, where

$$\Delta_n = \max_{1\le i\le n} |t_i^n - t_{i-1}^n|.$$

Remark 2.2. We propose an asymptotically distribution free test based on the sampling scheme (2.2), namely, high frequency data. We note there is a huge literature on discrete time approximations of statistical estimators for diffusion processes; see, for example, the Introduction of Gobet et al. [4] for a review including not only high frequency but also low frequency data. In the context of our goodness-of-fit test, however, it seems difficult to obtain asymptotically distribution free results based on low frequency data. Our result for this problem is related to prior work [14] in which some Kolmogorov–Smirnov-type tests based on smoothing were considered. The ideal assertion for the Kolmogorov–Smirnov type tests is still an open problem because it needs a weak convergence theorem in $\ell^\infty(\mathbb{R})$.

Under the setting stated above, the problem is to conduct a goodness-of-fit test of (2.1), that is, we wish to test the null hypothesis $H_0 : S = S_0$ *versus* $H_1 : S \ne S_0$ *for a given* S_0. Let us introduce the test statistic

$$\mathcal{D}_n = \int_{\mathbb{R}} \frac{|U_n(x;S_0)|^2}{\Psi_{S_0,\sigma}(\infty)} \frac{\Psi_{S_0,\sigma}(dx)}{\Psi_{S_0,\sigma}(\infty)}, \tag{2.3}$$

where

$$x \rightsquigarrow U_n(x;S) = \sum_{i=1}^n 1_{(-\infty,x]}(X_{t_{i-1}^n}) \frac{X_{t_i^n} - X_{t_{i-1}^n} - S(X_{t_{i-1}^n})(t_i^n - t_{i-1}^n)}{\sqrt{t_n^n}},$$

and

$$\Psi_{S,\sigma}(x) = \int_{-\infty}^x \sigma(z)^2 \mu_{S,\sigma}(dz) \quad (x \in \mathbb{R}).$$

In the next subsection, we show that the asymptotic null distribution of $\mathcal{D}_n$ is

$$\int_0^1 |B(u)|^2 du. \tag{2.4}$$

2.2. Justification of proposed procedure

We justify our test procedure; denoting

$$m_i^n = \frac{\sigma(X_{t_{i-1}^n})(W_{t_i^n} - W_{t_{i-1}^n})}{\sqrt{t_n^n}} \quad (i = 1, \dots, n), \tag{2.5}$$
$$\tilde{m}_i^n = \frac{X_{t_i^n} - X_{t_{i-1}^n} - S(X_{t_{i-1}^n})(t_i^n - t_{i-1}^n)}{\sqrt{t_n^n}} \quad (i = 1, \dots, n),$$

we suppose H_0 is true. Then, as argued in the proof of Proposition 2.1, the sequence $\{\tilde{m}_i^n\}_{i=1}^n$ is close to $\{m_i^n\}_{i=1}^n$, which is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_{i-1}\}_{i=1}^\infty$, and hence Theorem 1.1 yields the weak convergence in $L^2(\mathbb{R}, \Psi_{S_0,\sigma})$ of

$$x \rightsquigarrow \sum_{i=1}^n 1_{(-\infty,x]}(X_{t_{i-1}^n}) m_i^n,$$

which we denote for simplicity by $M_n^b(x)$.

Proposition 2.1. *Let ν be any finite Borel measure on $\mathbb{R}$. Assume* **(A1)** *and* **(A2)***. Then, $x \rightsquigarrow U_n(x; S)$ converges weakly in $L^2(\mathbb{R}, \nu)$ to $x \rightsquigarrow B \circ \Psi_{S,\sigma}(x)$ as $n \to \infty$ with (2.2), where $B(\cdot)$ is a standard Brownian motion and*

$$\Psi_{S,\sigma}(x) = \int_{-\infty}^x \sigma(z)^2 \mu_{S,\sigma}(dz) \quad (x \in \mathbb{R}).$$

Proof. Define

$$x \rightsquigarrow M_n^a(x) = \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^n 1_{(-\infty,x]}(X_{t_{i-1}^n}) \int_{t_{i-1}^n}^{t_i^n} \sigma(X_s) dW_s,$$

and

$$x \rightsquigarrow M_n^b(x) = \sum_{i=1}^n 1_{(-\infty,x]}(X_{t_{i-1}^n}) m_i^n,$$

where $\{m_i^n\}_{i=1}^n$ is defined in (2.5).

Straightforwardly from (2.2), $|U_n(\cdot; S) - M_n^a(\cdot)|$ converges in probability to zero under the uniform metric, and therefore under the $L^2(\mathbb{R}, \nu)$-metric.

We next show that $M_n^a(\cdot) - M_n^b(\cdot)$ converges weakly in $L^2(\mathbb{R}, \nu)$ to zero (the degenerate random field) and that $M_n^b(\cdot)$ converges to $B \circ \Psi_{S,\sigma}(\cdot)$; the assertion of the lemma then follows from Slutsky's lemma. To show these two weak convergence claims, we apply Theorem 1.1 for

$$x \rightsquigarrow \sum_{i=1}^n \xi_i^n(x)$$

with

$$\xi_i^n(x) = \frac{1}{\sqrt{t_n^n}} 1_{(-\infty,x]}(X_{t_{i-1}^n}) \int_{t_{i-1}^n}^{t_i^n} (\sigma(X_s) - \sigma(X_{t_{i-1}^n}))dW_s \quad (i = 1, \dots, n) \tag{2.6}$$

and

$$\xi_i^n(x) = 1_{(-\infty,x]}(X_{t_{i-1}^n}) m_i^n, \quad (i = 1, \dots, n) \tag{2.7}$$

respectively. In Assumption 1.1, Condition (i) for (2.6) in which the limit is zero is clear, whereas that for (2.7) is proved using Lemma 2.1 (iii). Condition (ii) is indeed satisfied. Conditions (iii) and (iv) are immediate from stationarity (as for (2.6); we use also the latter inequality of Lemma 2.1 (i)). This completes the proof. □

The random variable of the limit satisfies

$$\int_{\mathbb{R}} \frac{|B(\Psi_{S,\sigma}(x))|^2}{\Psi_{S,\sigma}(\infty)} \frac{\Psi_{S,\sigma}(dx)}{\Psi_{S,\sigma}(\infty)} =^d \int_0^1 |B(u)|^2 du,$$

where the notation $=^d$ means the distributions are the same. Hence, applying the continuous mapping theorem, we obtain the following corollary.

Corollary 2.1. *Suppose that* **(A1)** *and* **(A2)** *are satisfied for a given, specific* S_0 *and* σ. *If* H_0 *is true, then* $\mathcal{D}_n$ *converges in distribution to* (2.4) *as* $n \to \infty$ *with* (2.2).

We close this subsection with a remark on the consistency of the test. Let us write H_1 as

$$\int_{\mathbb{R}} \left| \int_{-\infty}^{x} \{S_0(z) - S(z)\} \mu_{S,\sigma}(dz) \right|^2 \Psi_{S_0,\sigma}(dx) > 0. \tag{2.8}$$

Hereafter, (2.8) is assumed to be true. Observe that

$$\begin{aligned}
& \Psi_{S_0,\sigma}(\infty)\mathcal{D}_n^{1/2} \\
= \;& \left(\int_{\mathbb{R}} |U_n(x;S_0)|^2 \Psi_{S_0,\sigma}(dx) \right)^{1/2} \\
\geq \;& \sqrt{t_n^n} \left(\int_{\mathbb{R}} |H_n(x)|^2 \, \Psi_{S_0,\sigma}(dx) \right)^{1/2} - \left(\int_{\mathbb{R}} |U_n(x;S)|^2 \Psi_{S_0,\sigma}(dx) \right)^{1/2},
\end{aligned}$$

where

$$H_n(\cdot) = \frac{1}{t_n^n} \sum_{i=1}^{n} 1_{(-\infty,\cdot]}(X_{t_{i-1}^n}) \{S_0(X_{t_{i-1}^n}) - S(X_{t_{i-1}^n})\} |t_i^n - t_{i-1}^n|.$$

Applying Proposition 2.1 and the continuous mapping theorem, the second term of the right-hand side is $O_P(1)$. To prove that the probability that the first term is bounded by M tends to zero as $n \to \infty$ for any $M > 0$, we first note that

$$H_n(x) \to^p \int_{-\infty}^{x} \{S_0(z) - S(z)\} \mu_{S,\sigma}(dz)$$

for all $x \in \mathbb{R}$; this follows from Lemma 2.1 (iii) presented in the next subsection. Showing that this convergence holds uniformly in x is straightforward. Hence

$$\int_{\mathbb{R}} |H_n(x)|^2 \,\Psi_{S_0,\sigma}(dx) \to^p \int_{\mathbb{R}} \left| \int_{-\infty}^{x} \{S_0(z) - S(z)\} \mu_{S,\sigma}(dz) \right|^2 \Psi_{S_0,\sigma}(dx) > 0.$$

Therefore, $P(\mathcal{D}_n > M) = P(\Psi_{S_0,\sigma}(\infty)\mathcal{D}_n^{1/2} > \Psi_{S_0,\sigma}(\infty)M^{1/2}) \to 1$ holds for any constant $M > 0$.

2.3. A technical lemma

We next prove a lemma that we have already used.

Lemma 2.1. *Let X be a solution of the SDE (2.1) with (S,σ) satisfying* **(A1)***. Let p be a positive integer, and assume $\sup_{t\in[0,\infty)} E|X_t|^p < \infty$.*

(i) There exists a constant $C_{p,S,\sigma} > 0$ depending only on p, (S,σ) such that if $|t_i - t_{i-1}| \le 1$ then

$$E\left[\sup_{s\in[t_{i-1}^n, t_i^n]} |X_s - X_{t_{i-1}^n}|^p \middle| \mathcal{F}_{t_{i-1}^n} \right] \le C_{p,S,\sigma} |t_i^n - t_{i-1}^n|^{p/2} (1 + |X_{t_{i-1}^n}|)^p,$$

$$E\left[\sup_{s\in[t_{i-1}^n, t_i^n]} |X_s|^p \middle| \mathcal{F}_{t_{i-1}^n} \right] \le C_{p,S,\sigma} (1 + |X_{t_{i-1}^n}|)^p.$$

(ii) Given p Lipschitz continuous functions $\mathbf{g} = (g_1, ..., g_p)$, there exists a constant $C_{p,\mathbf{g},S,\sigma} > 0$ depending also on (S,σ) such that if $|t_i^n - t_{i-1}^n| \le 1$ then

$$\begin{aligned} & E\left[\sup_{s\in[t_{i-1}^n, t_i^n]} \left| \prod_{j=1}^{p} g_j(X_s) - \prod_{j=1}^{p} g_j(X_{t_{i-1}^n}) \right| |\mathcal{F}_{t_{i-1}^n} \right] \\ \le \quad & C_{p,\mathbf{g},S,\sigma} |t_i^n - t_{i-1}^n|^{1/2} (1 + |X_{t_{i-1}^n}|)^p. \end{aligned}$$

(iii) Assume that X is ergodic with absolutely continuous invariant distribution μ. Given $x \in \mathbb{R}$ and $p-1$ Lipschitz continuous functions $\mathbf{g} = (g_1, ..., g_{p-1})$ such that $\prod_{j=1}^{p-1} g_j$ is μ-integrable, then with $\Delta_n \to 0$ it holds that

$$\frac{1}{t_n^n} \sum_{i=1}^{n} 1_{(-\infty,x]}(X_{t_{i-1}^n}) \prod_{j=1}^{p-1} g_j(X_{t_{i-1}^n}) |t_i^n - t_{i-1}^n| \to^p \int_{-\infty}^{x} \prod_{j=1}^{p-1} g_j(z) \mu(dz).$$

(This assertion is true also for $p = 1$ if we read $\prod_{j=1}^{1-1} g_j \equiv 1$.)

Proof. The assertion (i) is well-known; see, for example, Kessler [7]. The assertion (ii) can be proven using (i). This leaves the assertion (iii). Writing $g(z) = \prod_{j=1}^{p-1} g_j(z)$, we may assume that all g_j's are nonnegative without loss of generality. (For the general case, notice that g is represented as a sum of terms of the form $a\prod_{j=1}^{p-1}\widetilde{g}_j$ where $\widetilde{g}_j = g_j \vee 0$ or $(-g_j)\vee 0$ and $a = 1$ or -1.) For any $\varepsilon > 0$, choose two Lipschitz continuous functions l, u such that $l \le 1_{(-\infty,x]} \le u$ and that $\int_{\mathbb{R}} |u(z) - l(z)|g(z)\mu(dz) < \varepsilon$. Then it holds that

$$\begin{aligned}
& \frac{1}{t_n^n}\sum_{i=1}^n 1_{(-\infty,x]}(X_{t_{i-1}^n})g(X_{t_{i-1}^n})|t_i^n - t_{i-1}^n| \\
\le \quad & \frac{1}{t_n^n}\sum_{i=1}^n u(X_{t_{i-1}^n})g(X_{t_{i-1}^n})|t_i^n - t_{i-1}^n| \\
= \quad & \frac{1}{t_n^n}\int_0^{t_n^n} u(X_s)g(X_s)ds + O_P(\Delta_n^{1/2}) \\
\to^p \quad & \int_{\mathbb{R}} u(z)g(z)\mu(dz) \\
\le \quad & \int_{-\infty}^x g(z)\mu(dz) + \varepsilon.
\end{aligned}$$

By the same argument, replacing u by l, we finally get

$$\left| \frac{1}{t_n^n}\sum_{i=1}^n 1_{(-\infty,x]}(X_{t_{i-1}^n})g(X_{t_{i-1}^n})|t_i^n - t_{i-1}^n| - \int_{-\infty}^x g(z)\mu(dz) \right| \le \varepsilon + o_P(1).$$

Because the choice of ε is arbitrary, we have proven the assertion of (iii). This completes the proof. $\square$

3. Anderson–Darling-type goodness-of-fit test for nonlinear time series

In this section, we consider the goodness-of-fit test procedure for a nonlinear time series model. The procedure is justified using Theorem 1.2.

3.1. Problem setting and test procedure

We consider a strictly stationary ergodic stochastic process $\{X_i\}_{i=-\infty}^{\infty}$ given by

$$X_i = S(X_{i-1}) + \sigma(X_{i-1})\varepsilon_i \quad (i \in \mathbb{Z}), \tag{3.1}$$

where $S(\cdot)$ is a measurable function, $\sigma(\cdot)$ is a known measurable function satisfying $\inf_{x\in\mathbb{R}}\sigma(x) > 0$, and $\{\varepsilon_i\}_{i=-\infty}^{\infty}$ is an unobserved sequence of absolutely continuous random variables satisfying

$$P(\varepsilon_i \le 0|\mathcal{F}_{i-1}) = 1/2 \quad \text{a.s.},$$

where $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$ is the filtration defined by $\mathcal{F}_i = \sigma(X_j : j \le i)$ for all $i \in \mathbb{Z}$. In this section, no moment condition on ε_i $(i \in \mathbb{Z})$ is assumed.

We introduce the following assumption on $S(\cdot)$ and $\sigma(\cdot)$.

(B) The process $\{X_i\}_{i=-\infty}^{\infty}$ is stationary and ergodic with absolutely continuous invariant law $\mu_{S,\sigma}$, where ergodicity is in the sense of almost sure convergence, that is,

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \to \int_{\mathbb{R}} g(x)\mu_{S,\sigma}(dx) \quad \text{a.s.}$$

for every $\mu_{S,\sigma}$-integrable function $g(\cdot)$. Moreover, the distribution function $\Psi_{S,\sigma}$ of $\mu_{S,\sigma}$ satisfies

$$\int_{\mathbb{R}} \frac{\mu_{S,\sigma}(dx)}{\sqrt{\Psi_{S,\sigma}(x)}} < \infty.$$

In our problem, a time series $\{X_i\}_{i=0}^{n}$ is observed from the stochastic process (3.1).

Under the setting above, the problem is to conduct a goodness-of-fit test of (3.1), that is, we wish to test the null hypothesis $H_0 : S = S_0$ *versus* $H_1 : S \neq S_0$ *for a given* S_0. Let us define the test statistic

$$\mathcal{T}_n = \int_{\mathbb{R}} \frac{1}{n\Psi_{S_0,\sigma}(x)} \left(\sum_{i=1}^{n} \operatorname{sign}(X_i - S_0(X_{i-1})) 1_{(-\infty,x]}(X_{i-1}) \right)^2 \mu_{S_0,\sigma}(dx), \tag{3.2}$$

where $\operatorname{sign}(\cdot) = -1_{(-\infty,0)}(\cdot) + 1_{(0,\infty)}(\cdot)$. As is shown in the next subsection, the asymptotic null distribution of $\mathcal{T}_n$ is

$$\int_0^1 \frac{|B(u)|^2}{u} du. \tag{3.3}$$

Remark 3.1. Our statistic contains $\operatorname{sign}(\cdot)$ along the lines of Erlenmaier [2] and Section 7.3 of Nishiyama [17]. Of course, if the corresponding required condition on $\{\varepsilon_i\}_{i=1}^{\infty}$ is satisfied, other functions mentioned by Koul and Stute [10], can be used. Some examples are $f(\cdot) = \cdot$, $f(\cdot) = 1_{(0,\infty)}(\cdot) - (1-\alpha)$, and other bounded functions. Note that Nishiyama [18] considered another statistic similar to that used by Masuda et al. [14]. A merit of $f(\cdot) = \operatorname{sign}(\cdot)$ is its robustness against outliers.

Remark 3.2. Our procedure can be regarded as an Anderson–Darling-type statistic in the sense that

$$E\left[\int_0^1 \frac{|B(u)|^2}{u} du\right] = 1.$$

3.2. Justification of proposed procedure

We justify our test procedure using Theorem 1.2. Let

$$x \rightsquigarrow \xi_i^n(x) = \frac{1}{\sqrt{\Psi_{S_0,\sigma}(x)}} 1_{(-\infty,x]}(X_{i-1}) m_i^n \quad (i = 1, \ldots, n),$$

where

$$m_i^n = \frac{\operatorname{sign}(X_i - S_0(X_{i-1}))}{\sqrt{n}} \quad (i = 1, \ldots, n).$$

Suppose that H_0 is true. Then $\{m_i^n\}_{i=1}^n$ is a martingale difference sequence with respect to $\{\mathcal{F}_i\}_{i=0}^n$ for which

$$(m_i^n)^2 = \frac{1}{n} \quad \text{a.s.} \quad (i = 1, \ldots, n)$$

holds. We use Theorem 1.2 with $w = (\Psi_{S_0,\sigma})^{-1/2}$. From stationarity and ergodicity, Assumption 1.2 can be checked. Indeed, we find

$$E^n\left[\frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x]}(X_{i-1}^n)\right] = E\left[\frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x]}(\zeta)\right] = \Psi_{S_0,\sigma}(x)$$

holds; here, ζ is a random variable following $\mu_{S_0,\sigma}$. Hence, if **(B)** is satisfied, we are then able to check Assumption 1.2 by setting $\Psi = \Phi = \Lambda = \Psi_{S_0,\sigma}$ and $\delta = 1$. We thus have

$$\sum_{i=1}^n \xi_i^n \Rightarrow \frac{B \circ \Psi_{S_0,\sigma}}{\sqrt{\Psi_{S_0,\sigma}}} \quad \text{in} \quad L^2(\mathbb{R}, \nu)$$

as $n \to \infty$ for any finite Borel measure ν. Therefore, the continuous mapping theorem and

$$\int_{\mathbb{R}} \frac{|B(\Psi_{S_0,\sigma}(x))|^2}{\Psi_{S_0,\sigma}(x)} \mu_{S_0,\sigma}(dx) = \int_{\mathbb{R}} \frac{|B(\Psi_{S_0,\sigma}(x))|^2}{\Psi_{S_0,\sigma}(x)} \frac{\mu_{S_0,\sigma}(dx)}{\Psi_{S_0,\sigma}(\infty)} =^d \int_0^1 \frac{|B(u)|^2}{u} du$$

yield the following assertion.

Proposition 3.1. *Suppose that **(B)** is satisfied for a given, specific S_0 and σ. If H_0 is true, then $\mathcal{T}_n$ defined in (3.2) converges in distribution to (3.3) as $n \to \infty$.*

Remark 3.3. In this paper, we do not demonstrate the weak convergence of

$$x \rightsquigarrow \sum_{i=1}^n \frac{1}{\sqrt{\hat{\Psi}_n(x)}} 1_{(-\infty,x]}(X_{i-1}) m_i^n$$

in $L^2(\mathbb{R}, \nu)$ for which

$$\hat{\Psi}_n(x) = \frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x]}(X_{i-1}^n).$$

Finally, we briefly discuss the consistency of the test. To privide a clear sufficient condition, we consider the case that $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables. Let F_ε be the distribution fuction of ε_1. Note that $F_\varepsilon(0) = 1/2$ is assumed. Then, for all i it holds that

$$P(S(X_{i-1}) - S_0(X_{i-1}) + \sigma(X_{i-1})\varepsilon_i \le 0 | \mathcal{F}_{i-1})$$

$$
\begin{aligned}
&= P\left(\varepsilon_i \le \frac{S_0(X_{i-1}) - S(X_{i-1})}{\sigma(X_{i-1})} | \mathcal{F}_{i-1}\right) \\
&= F_\varepsilon\left(\frac{S_0(X_{i-1}) - S(X_{i-1})}{\sigma(X_{i-1})}\right) \quad \text{a.s.}
\end{aligned}
$$

Define a measurable function δ by

$$
\delta(x) = \frac{1}{2} - F_\varepsilon\left(\frac{S_0(x) - S(x)}{\sigma(x)}\right) \quad (x \in \mathbb{R}).
$$

Note that $\delta = 0$ if $H_0 : S = S_0$ is true. Let us write H_1 as

$$
\int_{\mathbb{R}} \left| \int_{-\infty}^{x} \delta(z) \mu_{S,\sigma}(dz) \right|^2 \mu_{S_0,\sigma}(dx) > 0,
$$

and assume it to be true thereafter. From $\Psi_{S_0,\sigma}(x) \le \Psi_{S_0,\sigma}(\infty) = 1$, it follows that

$$
\mathcal{T}_n^{1/2} \ge \left\{ \int_{\mathbb{R}} \frac{1}{n} \left(\sum_{i=1}^{n} \operatorname{sign}(X_i - S_0(X_{i-1})) 1_{(-\infty,x]}(X_{i-1}) \right)^2 \mu_{S_0,\sigma}(dx) \right\}^{1/2}.
$$

Here, the right-hand side is bounded below by

$$
\begin{aligned}
&\sqrt{n} \times \left\{ 4 \int_{\mathbb{R}} \left(\frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,x]}(X_{i-1}) \delta(X_{i-1}) \right)^2 \mu_{S_0,\sigma}(dx) \right\}^{1/2} \\
&- \left\{ \int_{\mathbb{R}} \left(\sum_{i=1}^{n} 1_{(-\infty,x]}(X_{i-1}) \check{m}_i^n \right)^2 \mu_{S_0,\sigma}(dx) \right\}^{1/2},
\end{aligned}
$$

where

$$
\check{m}_i^n = \frac{1}{\sqrt{n}} \left(\operatorname{sign}(X_i - S_0(X_{i-1})) - 2\delta(X_{i-1}) \right) \quad (i = 1, \ldots, n).
$$

The first term tends to positive infinity in probability because

$$
\begin{aligned}
&\int_{\mathbb{R}} \left(\frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,x]}(X_{i-1}) \delta(X_{i-1}) \right)^2 \mu_{S_0,\sigma}(dx) \\
&\to^p \int_{\mathbb{R}} \left| \int_{-\infty}^{x} \delta(z) \mu_{S,\sigma}(dz) \right|^2 \mu_{S_0,\sigma}(dx) > 0
\end{aligned}
$$

which follows from ergodicity, whereas the second term is $O_P(1)$ which is a consequence of Theorem 1.2 because $\{\check{m}_i^n\}_{i=1}^{\infty}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_i\}_{i=0}^{\infty}$. Therefore, $P(\mathcal{T}_n > M) = P(\mathcal{T}_n^{1/2} > M^{1/2}) \to 1$ holds for any constant $M > 0$.

Appendix A: Proof of Theorem 1.1

By Prohorov's tightness criterion for Hilbert space valued random sequences (see, e.g., Theorem 1.8.4 by van der Vaart and Wellner [19]), it suffices to show the following two lemmas.

Lemma A.1. *Under Assumptions 1.1 (i) and (ii), $\langle Z_n, h\rangle$ converges in distribution to $\langle G, h\rangle$ for any $h \in L^2(\mathbb{R}, \nu)$ as $n \to \infty$.*

Lemma A.2. *Under Assumptions 1.1 (iii) and (iv), it holds that*

$$\lim_{J\to\infty} \limsup_{n\to\infty} E^n\left[\sum_{j=J}^{\infty} \langle Z_n, e_j\rangle^2\right] = 0.$$

A.1. Proof of Lemma A.1

Because

$$\begin{aligned}\langle Z_n^w, h\rangle &= \int_{\mathbb{R}} \sum_{i=1}^n 1_{(-\infty,x]}(X_{i-1}^n) m_i^n h(x)\nu(dx) \\ &= \sum_{i=1}^n \left(\int_{\mathbb{R}} 1_{(-\infty,x]}(X_{i-1}^n) h(x)\nu(dx)\right) m_i^n,\end{aligned}$$

we apply the martingale central limit theorem (Hall and Heyde [6], Corollary 3.1) for the martingale difference sequence

$$\left\{\left(\int_{\mathbb{R}} 1_{(-\infty,x]}(X_{i-1}^n) h(x)\nu(dx)\right) m_i^n\right\}_{i=1}^n.$$

It is straightforward to prove that Assumption 1.1 (i) leads to

$$\sup_{x\in\mathbb{R}} |R_n(x)| \to^p 0,$$

where

$$R_n(x) = \sum_{i=1}^n 1_{(-\infty,x]}(X_{i-1}^n) E^n[|m_i^n|^2 | \mathcal{F}_{i-1}^n] - \Psi(x). \tag{A.1}$$

Hence

$$\begin{aligned}&\sum_{i=1}^n \left(\int_{\mathbb{R}} 1_{(-\infty,x]}(X_{i-1}^n) h(x)\nu(dx)\right)^2 E^n[|m_i^n|^2 | \mathcal{F}_{i-1}^n] \\ =\ & \sum_{i=1}^n \int_{\mathbb{R}}\int_{\mathbb{R}} 1_{(-\infty,x\wedge y]}(X_{i-1}^n) h(x)h(y)\nu(dx)\nu(dy) E^n[|m_i^n|^2 | \mathcal{F}_{i-1}^n] \\ =\ & \int_{\mathbb{R}}\int_{\mathbb{R}} \Psi(x\wedge y) h(x)h(y)\nu(dx)\nu(dy)\end{aligned}$$

$$
\begin{aligned}
&+\int_{\mathbb{R}}\int_{\mathbb{R}} R_n(x\wedge y)h(x)h(y)\nu(dx)\nu(dy)\\
\to^p\quad &\int_{\mathbb{R}}\int_{\mathbb{R}} \Psi(x\wedge y)h(x)h(y)\nu(dx)\nu(dy).
\end{aligned}
$$

Moreover, we have

$$
E[\langle G,h\rangle^2]=\int_{\mathbb{R}}\int_{\mathbb{R}} \Psi(x\wedge y)h(x)h(y)\nu(dx)\nu(dy).
$$

What remains to be shown is the Lyapunov-type condition,

$$
\sum_{i=1}^n E^n\left[\left|\int_{\mathbb{R}} 1_{(-\infty,x]}(X_{i-1}^n)h(x)\nu(dx)\right|^{2+\delta}|m_i^n|^{2+\delta}|\mathcal{F}_{i-1}^n\right]\to^p 0. \tag{A.2}
$$

From

$$
\begin{aligned}
\left|\int_{\mathbb{R}} 1_{(-\infty,x]}(X_{i-1}^n)h(x)\nu(dx)\right|^{2+\delta} &\le \left(\int_{\mathbb{R}} 1_{(-\infty,x]}(X_{i-1}^n)\nu(dx)\right)^{\frac{2+\delta}{2}}\|h\|^{2+\delta}\\
&\le (\nu(\mathbb{R}))^{\frac{2+\delta}{2}}\|h\|^{2+\delta},
\end{aligned}
$$

the left-hand side of (A.2) is bounded above by

$$
(\nu(\mathbb{R}))^{\frac{2+\delta}{2}}\|h\|^{2+\delta}\sum_{i=1}^n E^n\left[|m_i^n|^{2+\delta}|\mathcal{F}_{i-1}^n\right],
$$

which converges in probability to 0 by Assumption 1.1 (ii).

This completes the proof. □

A.2. Proof of Lemma A.2

For simplicity, let us denote

$$
\xi_i^n(x)=1_{(-\infty,x]}(X_{i-1}^n)m_i^n\quad (x\in\mathbb{R}).
$$

It follows from Assumptions 1.1 (iii) and (iv) that

$$
\begin{aligned}
E^n[\langle\xi_i^n,e_j\rangle^2|\mathcal{F}_{i-1}^n] &= \langle 1_{(-\infty,\cdot]}(X_{i-1}^n),e_j\rangle^2 E^n[|m_i^n|^2|\mathcal{F}_{i-1}^n]\\
&\le \langle 1_{(-\infty,\cdot]}(X_{i-1}^n),e_j\rangle^2 c_i^n\phi(X_{i-1}^n)\\
&= c_i^n\left\langle 1_{(-\infty,\cdot]}(X_{i-1}^n)\sqrt{\phi(X_{i-1}^n)},e_j\right\rangle^2,
\end{aligned}
$$

which yields

$$
\sum_{i=1}^n E^n\left[\langle\xi_i^n,e_j\rangle^2\right]\le\sum_{i=1}^n c_i^n E\left[\langle\tilde{\xi},e_j\rangle^2\right]\le E\left[\langle\eta,e_j\rangle^2\right],
$$

where $x \rightsquigarrow \tilde{\xi}(x)$ is a $L^2(\mathbb{R}, \nu)$-valued random variable which follows the same distribution as $1_{(-\infty,x]}(\zeta)\sqrt{\phi(\zeta)}$, and

$$\eta = \left(\sup_n \sum_{i=1}^n c_i^n \right)^{1/2} \tilde{\xi}.$$

Hence,

$$\begin{aligned} E^n\left[\sum_{j=J}^\infty \left\langle \sum_{i=1}^n \xi_i^n, e_j \right\rangle^2 \right] &= \sum_{j=J}^\infty E^n\left[\left\langle \sum_{i=1}^n \xi_i^n, e_j \right\rangle^2 \right] \\ &= \sum_{j=J}^\infty \sum_{i=1}^n E^n\left[\langle \xi_i^n, e_j \rangle^2 \right] \\ &\leq \sum_{j=J}^\infty E\left[\langle \eta, e_j \rangle^2 \right] \\ &= E\left[\sum_{j=J}^\infty \langle \eta, e_j \rangle^2 \right]. \end{aligned}$$

Because $E[\|\eta\|^2] < \infty$, the dominated convergence theorem yields

$$\lim_{J\to\infty} E\left[\sum_{j=J}^\infty \langle \eta, e_j \rangle^2 \right] = E\left[\lim_{J\to\infty} \sum_{j=J}^\infty \langle \eta, e_j \rangle^2 \right] = 0.$$

This completes the proof. □

Appendix B: Proof of Theorem 1.2

Applying Prohorov's criterion, it suffices to show the following two lemmas.

Lemma B.1. *Under Assumption 1.2, $\langle Z_n^w, h \rangle$ converges in distribution to $\langle G^w, h \rangle$ for any $h \in L^2(\mathbb{R}, \nu)$ as $n \to \infty$.*

Lemma B.2. *Under Assumption 1.2 (i), it holds that*

$$\lim_{J\to\infty} \limsup_{n\to\infty} E^n \left[\sum_{j=J}^\infty \langle Z_n^w, e_j \rangle^2 \right] = 0.$$

In proving these lemmas, let n be sufficiently large such that (1.5) and (1.6) hold.

B.1. Proof of Lemma B.1

Because

$$\begin{aligned}\langle Z_n^w, h\rangle &= \int_{\mathbb{R}} \sum_{i=1}^n w(x) 1_{(-\infty,x]}(X_{i-1}^n) m_i^n h(x)\nu(dx) \\ &= \sum_{i=1}^n \left(\int_{\mathbb{R}} w(x) 1_{(-\infty,x]}(X_{i-1}^n) h(x)\nu(dx)\right) m_i^n,\end{aligned}$$

we apply the martingale central limit theorem for the martingale difference sequence

$$\left\{\left(\int_{\mathbb{R}} w(x) 1_{(-\infty,x]}(X_{i-1}^n) h(x)\nu(dx)\right) m_i^n\right\}_{i=1}^n.$$

First we show that

$$\sum_{i=1}^n \left(\int_{\mathbb{R}} w(x) 1_{(-\infty,x]}(X_{i-1}^n) h(x)\nu(dx)\right)^2 E^n[(m_i^n)^2|\mathcal{F}_{i-1}^n] \tag{B.1}$$

converges in first mean to

$$E[\langle G^w, h\rangle^2] = \int_{\mathbb{R}}\int_{\mathbb{R}} \Psi(x\wedge y) w(x)w(y)h(x)h(y)\nu(dx)\nu(dy).$$

Note that (B.1) equals

$$\begin{aligned}&\sum_{i=1}^n \left(\int_{\mathbb{R}}\int_{\mathbb{R}} w(x)w(y) 1_{(-\infty,x\wedge y]}(X_{i-1}^n) h(x)h(y)\nu(dx)\nu(dy)\right) E^n[(m_i^n)^2|\mathcal{F}_{i-1}^n] \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \sum_{i=1}^n 1_{(-\infty,x\wedge y]}(X_{i-1}^n) E^n[(m_i^n)^2|\mathcal{F}_{i-1}^n] w(x)w(y)h(x)h(y)\nu(dx)\nu(dy).\end{aligned}$$

We use the dominated convergence theorem to see

$$\int_{\mathbb{R}}\int_{\mathbb{R}} E^n\left[|R_n(x\wedge y)|\right] w(x)w(y)|h(x)h(y)|\nu(dx)\nu(dy) \to 0, \tag{B.2}$$

where $R_n(\cdot)$ is defined in (A.1), because

$$\begin{aligned}&E^n\left[\left|\int_{\mathbb{R}}\int_{\mathbb{R}} R_n(x\wedge y) w(x)w(y)h(x)h(y)\nu(dx)\nu(dy)\right|\right] \\ &\le \int_{\mathbb{R}}\int_{\mathbb{R}} E^n\left[|R_n(x\wedge y)|\right] w(x)w(y)|h(x)h(y)|\nu(dx)\nu(dy). \qquad \text{(B.3)}\end{aligned}$$

From (1.4), for all x and y, we have

$$E^n\left[|R_n(x\wedge y)|\right] w(x)w(y)|h(x)h(y)| \to 0.$$

Moreover, for all x and y, we have

$$\begin{aligned}
& E^n\left[|R_n(x\wedge y)|\right] w(x)w(y)|h(x)h(y)| \\
\leq\ & \left(I_n(x\wedge y)+\Psi(x\wedge y)\right) w(x)w(y)|h(x)h(y)| \\
\leq\ & \left(\Psi(x\wedge y)+\Phi(x\wedge y)\right) w(x)w(y)|h(x)h(y)| \\
\leq\ & \left(\sqrt{\Psi(x)\Psi(y)}+\sqrt{\Phi(x)\Phi(y)}\right) w(x)w(y)|h(x)h(y)| \\
=\ & \sqrt{\Psi(x)\Psi(y)}w(x)w(y)|h(x)h(y)|+\sqrt{\Phi(x)\Phi(y)}w(x)w(y)|h(x)h(y)|,
\end{aligned}$$

where we have used

$$\begin{aligned}
E^n[|R_n(x\wedge y)|] &\leq E^n\left[\sum_{i=1}^n 1_{(-\infty,x\wedge y]}(X_{i-1}^n)E^n[(m_i^n)^2|\mathcal{F}_{i-1}^n]\right]+\Psi(x\wedge y) \\
&= I_n(x\wedge y)+\Psi(x\wedge y)
\end{aligned}$$

and $\Psi(x\wedge y)\leq\sqrt{\Psi(x)\Psi(y)}$ and $\Phi(x\wedge y)\leq\sqrt{\Phi(x)\Phi(y)}$ which follow from the monotonicity of Ψ and Φ. Furthermore, the Cauchy–Schwarz inequality yields

$$\begin{aligned}
& \int_{\mathbb{R}}\int_{\mathbb{R}}\sqrt{\Psi(x)\Psi(y)}w(x)w(y)|h(x)h(y)|\nu(dx)\nu(dy) \\
\leq\ & \left(\int_{\mathbb{R}}\int_{\mathbb{R}}\Psi(x)\Psi(y)(w(x))^2(w(y))^2\nu(dx)\nu(dy)\right)^{1/2} \\
& \left(\int_{\mathbb{R}}\int_{\mathbb{R}}|h(x)h(y)|^2\nu(dx)\nu(dy)\right)^{1/2} \\
=\ & \left\{\int_{\mathbb{R}}\Psi(x)(w(x))^2\nu(dx)\right\}\|h\|^2<\infty
\end{aligned}$$

and

$$\begin{aligned}
& \int_{\mathbb{R}}\int_{\mathbb{R}}\sqrt{\Phi(x)\Phi(y)}w(x)w(y)|h(x)h(y)|\nu(dx)\nu(dy) \\
\leq\ & \left\{\int_{\mathbb{R}}\Phi(x)(w(x))^2\nu(dx)\right\}\|h\|^2<\infty.
\end{aligned}$$

Therefore, the dominated convergence theorem implies (B.2).

Next, we show the Lyapunov-type condition; specifically, the nonnegative valued random variable

$$\sum_{i=1}^n E^n\left[\left|\int_{\mathbb{R}} w(x)1_{(-\infty,x]}(X_{i-1}^n)h(x)\nu(dx)\right|^{2+\delta}|m_i^n|^{2+\delta}|\mathcal{F}_{i-1}^n\right]$$

converges in probability to 0. Because the Cauchy–Schwarz inequality yields

$$\left|\int_{\mathbb{R}} w(x)1_{(-\infty,x]}(X_{i-1}^n)h(x)\nu(dx)\right|^{2+\delta}$$

$$\leq \quad \left(\int_{\mathbb{R}} (w(x))^2 1_{(-\infty,x]}(X_{i-1}^n) \nu(dx) \right)^{\frac{2+\delta}{2}} \|h\|^{2+\delta},$$

it suffices to show the convergence of

$$\sum_{i=1}^n E^n \left[\left(\int_{\mathbb{R}} (w(x))^2 1_{(-\infty,x]}(X_{i-1}^n) \nu(dx) \right)^{\frac{2+\delta}{2}} |m_i^n|^{2+\delta} \right]$$

to 0. Moreover, this equation can be evaluated by

$$\begin{aligned} & E^n \left[\sum_{i=1}^n \left(\int_{\mathbb{R}} (w(x))^2 1_{(-\infty,x]}(X_{i-1}^n) \nu(dx) \right)^{\frac{2+\delta}{2}} |m_i^n|^{2+\delta} \right] \\ \leq \quad & (\nu(\mathbb{R}))^{\frac{\delta}{2}} E^n \left[\sum_{i=1}^n \int_{\mathbb{R}} 1_{(-\infty,x]}(X_{i-1}^n)(w(x))^{2+\delta} \nu(dx) |m_i^n|^{2+\delta} \right] \\ = \quad & (\nu(\mathbb{R}))^{\frac{\delta}{2}} \int_{\mathbb{R}} \sum_{i=1}^n E^n \left[1_{(-\infty,x]}(X_{i-1}^n) |m_i^n|^{2+\delta} \right] (w(x))^{2+\delta} \nu(dx). \end{aligned}$$

Applying the dominated convergence theorem, the right-hand side converges to 0. Indeed, as for the integrand, the convergence

$$\sum_{i=1}^n E^n \left[1_{(-\infty,x]}(X_{i-1}^n) |m_i^n|^{2+\delta} \right] (w(x))^{2+\delta} \to 0$$

holds for all x, and

$$\sum_{i=1}^n E^n \left[1_{(-\infty,x]}(X_{i-1}^n) |m_i^n|^{2+\delta} \right] (w(x))^{2+\delta} \leq \Lambda(x)(w(x))^{2+\delta},$$

the right-hand side of which is ν-integrable.

This completes the proof. □

B.2. Proof of Lemma B.2

In this subsection, for simplicity, let us denote

$$\xi_i^n(x) = w(x) 1_{(-\infty,x]}(X_{i-1}^n) m_i^n \quad (x \in \mathbb{R}).$$

It holds that

$$\begin{aligned} & E^n \left[\sum_{j=J}^{\infty} \left\langle \sum_{i=1}^n \xi_i^n, e_j \right\rangle^2 \right] \\ = \quad & E^n \left[\left\| \sum_{i=1}^n \xi_i^n \right\|^2 - \sum_{j=1}^J \left\langle \sum_{i=1}^n \xi_i^n, e_j \right\rangle^2 \right] \end{aligned}$$

$$= \quad E^n\left[\left\|\sum_{i=1}^n \xi_i^n\right\|^2\right] - \sum_{j=1}^J E^n\left[\left\langle \sum_{i=1}^n \xi_i^n, e_j\right\rangle^2\right]. \qquad \text{(B.4)}$$

Because

$$E^n[\langle \xi_i^n, \xi_j^n\rangle] = E^n[E^n[\langle \xi_i^n, \xi_j^n\rangle | \mathcal{F}_{j-1}^n]] = 0$$

for $i < j$, the first term in the right-hand side of (B.4) becomes

$$\begin{aligned} E^n\left[\left\|\sum_{i=1}^n \xi_i^n\right\|^2\right] &= E^n\left[\sum_{i=1}^n \|\xi_i^n\|^2\right] + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n E^n\left[\langle \xi_i^n, \xi_j^n\rangle\right] \\ &= E^n\left[\sum_{i=1}^n \|\xi_i^n\|^2\right]. \end{aligned}$$

Applying the dominated convergence theorem, we have

$$\begin{aligned} &\lim_{n\to\infty} E^n\left[\sum_{i=1}^n \|\xi_i^n\|^2\right] \\ = \; &\lim_{n\to\infty}\int_{\mathbb{R}} \sum_{i=1}^n E^n[1_{(-\infty,x]}(X_{i-1}^n)|m_i^n|^2](w(x))^2\nu(dx) \\ = \; &\lim_{n\to\infty}\int_{\mathbb{R}} I_n(x)(w(x))^2\nu(dx) \\ = \; &\int_{\mathbb{R}} \Psi(x)(w(x))^2\nu(dx) \\ = \; &\int_{\mathbb{R}} E\left[(B(\Psi(x)))^2\right](w(x))^2\nu(dx) \\ = \; &E[\|wB\circ\Psi\|^2], \end{aligned}$$

where $wB\circ\Psi$ means $w(\cdot)B(\Psi(\cdot))$. That is because for all $x \in \mathbb{R}$, we have $I_n(x)(w(x))^2 \to \Psi(x)(w(x))^2$ and $I_n(x)(w(x))^2 \le \Phi(x)(w(x))^2$, the right-hand side of which is ν-integrable.

For the second term on the right-hand side of (B.4), because $\{\langle \xi_i^n, e_j\rangle\}_{i=1}^n$ is a martingale difference sequence, we have

$$\begin{aligned} &E^n\left[\left\langle \sum_{i=1}^n \xi_i^n, e_j\right\rangle^2\right] \\ = \; &E^n\left[\left(\sum_{i=1}^n \langle \xi_i^n, e_j\rangle\right)^2\right] \\ = \; &\sum_{i=1}^n E^n\left[\langle \xi_i^n, e_j\rangle^2\right] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} E^n \left[\int_{\mathbb{R}} \int_{\mathbb{R}} w(x) w(y) 1_{(-\infty, x \wedge y]}(X_{i-1}^n) e_j(x) e_j(y) \nu(dx) \nu(dy) (m_i^n)^2 \right] \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} I_n(x \wedge y) w(x) w(y) e_j(x) e_j(y) \nu(dx) \nu(dy).
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\sum_{j=1}^{J} E^n \left[\left\langle \sum_{i=1}^{n} \xi_i^n, e_j \right\rangle^2 \right] \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \left(I_n(x \wedge y) w(x) w(y) \sum_{j=1}^{J} e_j(x) e_j(y) \right) \nu(dx) \nu(dy).
\end{aligned}
$$

The dominated convergence theorem yields

$$
\begin{aligned}
&\lim_{n \to \infty} \int_{\mathbb{R}} \int_{\mathbb{R}} \left(I_n(x \wedge y) w(x) w(y) \sum_{j=1}^{J} e_j(x) e_j(y) \right) \nu(dx) \nu(dy) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \lim_{n \to \infty} \left(I_n(x \wedge y) w(x) w(y) \sum_{j=1}^{J} e_j(x) e_j(y) \right) \nu(dx) \nu(dy). \quad \text{(B.5)}
\end{aligned}
$$

That is because, as for the integrand, we have

$$
\begin{aligned}
&\left| I_n(x \wedge y) w(x) w(y) \sum_{j=1}^{J} e_j(x) e_j(y) \right| \\
&\leq \Phi(x \wedge y) w(x) w(y) \sum_{j=1}^{J} |e_j(x) e_j(y)| \\
&\leq \sqrt{\Phi(x)\Phi(y)} w(x) w(y) \sum_{j=1}^{J} |e_j(x) e_j(y)|
\end{aligned}
$$

for all x and y, and

$$
\begin{aligned}
&\int_{\mathbb{R}} \int_{\mathbb{R}} \sqrt{\Phi(x)\Phi(y)} w(x) w(y) \sum_{j=1}^{J} |e_j(x) e_j(y)| \nu(dx) \nu(dy) \\
&= \sum_{j=1}^{J} \int_{\mathbb{R}} \int_{\mathbb{R}} \sqrt{\Phi(x)\Phi(y)} w(x) w(y) |e_j(x) e_j(y)| \nu(dx) \nu(dy) \\
&\leq \sum_{j=1}^{J} \left(\int_{\mathbb{R}} \int_{\mathbb{R}} |e_j(x) e_j(y)|^2 \nu(dx) \nu(dy) \right)^{1/2} \\
&\qquad \left(\int_{\mathbb{R}} \int_{\mathbb{R}} \Phi(x)\Phi(y) (w(x))^2 (w(y))^2 \nu(dx) \nu(dy) \right)^{1/2}
\end{aligned}
$$

$$= \quad J \int_{\mathbb{R}} \Phi(x)(w(x))^2 \nu(dx) < \infty.$$

Moreover, (B.5) equals

$$\begin{aligned}
&\int_{\mathbb{R}} \int_{\mathbb{R}} \Psi(x \wedge y) w(x) w(y) \sum_{j=1}^{J} e_j(x) e_j(y) \nu(dx) \nu(dy) \\
= \quad &\sum_{j=1}^{J} \int_{\mathbb{R}} \int_{\mathbb{R}} (\Psi(x) \wedge \Psi(y)) w(x) w(y) e_j(x) e_j(y) \nu(dx) \nu(dy).
\end{aligned}$$

In addition,

$$\begin{aligned}
&\sum_{j=1}^{J} E\left[\langle wB \circ \Psi, e_j \rangle^2\right] \\
= \quad &\sum_{j=1}^{J} E\left[\left(\int_{\mathbb{R}} B(\Psi(x)) w(x) e_j(x) \nu(dx)\right)^2\right] \\
= \quad &\sum_{j=1}^{J} \int_{\mathbb{R}} \int_{\mathbb{R}} E[B(\Psi(x)) B(\Psi(y))] w(x) w(y) e_j(x) e_j(y) \nu(dx) \nu(dy) \\
= \quad &\sum_{j=1}^{J} \int_{\mathbb{R}} \int_{\mathbb{R}} (\Psi(x) \wedge \Psi(y)) w(x) w(y) e_j(x) e_j(y) \nu(dx) \nu(dy).
\end{aligned}$$

From what have been already proven,

$$\lim_{J \to \infty} \limsup_{n \to \infty} E^n \left[\sum_{j=J}^{\infty} \left\langle \sum_{i=1}^{n} \xi_i^n, e_j \right\rangle^2\right]$$

equals

$$E\left[\|wB \circ \Psi\|^2\right] - \lim_{J \to \infty} E\left[\sum_{j=1}^{J} \langle wB \circ \Psi, e_j \rangle^2\right]. \tag{B.6}$$

Finally, applying the dominated convergence theorem, we have (B.6) equals

$$E\left[\|wB \circ \Psi\|^2\right] - E\left[\sum_{j=1}^{\infty} \langle wB \circ \Psi, e_j \rangle^2\right] = 0.$$

This completes the proof. □

Acknowledgements

The authors would like to express their sincere gratitude to an associate editor and an anonymous referee who provided us valuable comments. This study was partly supported by Japan Society for the Promotion of Science KAKENHI Grant Number 16H02791 (KT), 18K13454 (KT), 15K00062 (YN), and 18K11203 (YN). This study was mainly carried out when the first author was a member of Graduate School of Arts and Sciences, the University of Tokyo.

References

[1] Dachian, S., Kutoyants, Y.A. 2008. On the goodness-of-fit tests for some continuous time processes. *Statistical Models and Methods for Biomedical and Technical Systems*, 385–403, Stat. Ind. Technol., Birkhäuser Boston, Boston, MA. MR2462767

[2] Erlenmaier, U. 1997. A new criterion for tightness of stochastic processes and an application to Markov processes. *Preprint*.

[3] Escanciano, J.C. 2007. Weak convergence of non-stationary multivariate marked processes with applications to martingale testing. *J. Multivariate Anal.* **98**(7): 1321–1336. MR2364121

[4] Gobet, E., Hoffmann, M., Reiß, M. 2004. Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32**(5): 2223–2253. MR2102509

[5] González-Manteiga, W., Crujeiras, R.M. 2013. An updated review of goodness-of-fit tests for regression models. *TEST* **22**(3): 361–411. MR3093195

[6] Hall, P., Heyde, C.C. 1980. *Martingale Limit Theory and its Application.* Probability and Mathematical Statistics. Academic Press, Inc. MR0624435

[7] Kessler, M. 1997. Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**(2): 211–229. MR1455868

[8] Khmaladze, E.V. 1981. A martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26**: 240–257. MR0616619

[9] Kleptsyna, M., Kutoyants, Y.A. 2014. On asymptotically distribution free tests with parametric hypothesis for ergodic diffusion processes. *Stat. Inference Stoch. Process.* **17**(3): 295–319. MR3256842

[10] Koul, H.L., Stute, W. 1999. Nonparametric model checks for time series. *Ann. Statist.* **27**(1): 204–236. MR1701108

[11] Kutoyants, Y.A. 2004. *Statistical Inference for Ergodic Diffusion Processes*, Springer Series in Statistics. Springer-Verlag London, Ltd., London. MR2144185

[12] Kutoyants, Y.A. 2014. On asymptotic distribution of parameter free tests for ergodic diffusion processes. *Stat. Inference Stoch. Process.* **17**(2): 139–161. MR3219526

[13] Kutoyants, Y.A. 2016. On score-functions and goodness-of-fit tests for stochastic processes. *Math. Methods Statist.* **25**(2): 99–120. MR3519644

[14] Masuda, H., Negri, I., Nishiyama, Y. 2011. Goodness-of-fit test for ergodic diffusions by discrete-time observations: an innovation martingale approach. *J. Nonparametr. Stat.* **23**(2): 237–254. MR2801292

[15] Negri, I., Nishiyama, Y. 2009. Goodness of fit test for ergodic diffusion processes. *Ann. Inst. Statist. Math.* **61**(4): 919–928. MR2556771

[16] Negri, I., Nishiyama, Y. 2010. Review on goodness of fit tests for ergodic diffusion processes by different sampling schemes. *Economic Notes* **39**(1/2): 91–106. MR2556771

[17] Nishiyama, Y. 2000. *Entropy Methods for Martingales*. CWI Tract, **128**. Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam. MR1742612

[18] Nishiyama, Y. 2009. Goodness-of-fit test for a nonlinear time series. *J. Time Series Anal.* **30**(6): 674–681. MR2567660

[19] van der Vaart, A.W., Wellner, J.A. 1996. *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer Series in Statistics. Springer-Verlag, New York. MR1385671