Journal of Econometrics 169 (2012) 279–292

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Stationarity-based specification tests for diffusions when the process is nonstationary[☆]

Yacine Aït-Sahalia[a,b,*], Joon Y. Park[c,d,1]

[a] Department of Economics, Princeton University, Princeton, NJ 08544-1021, United States
[b] NBER, United States
[c] Department of Economics, Indiana University, Bloomington, IN 47405-7000, United States
[d] Sungkyunkwan University, Republic of Korea

ARTICLE INFO

Article history:
Available online 24 January 2012

ABSTRACT

We analyze in this paper the asymptotic behavior of the specification test of Aït-Sahalia (1996) for the stationary density of a diffusion process, but when the diffusion is not stationary. We consider integrated and explosive processes, as well as nearly integrated ones in the spirit of the local to unity analysis in classical unit root theory. We find that the behavior of the test predicted by the asymptotic distribution under an integrated process provides a better approximation to the small sample distribution of the test than that predicted by the asymptotic distribution under strict stationarity.

1. Introduction

Consider a given parametrization for a diffusion

$$dX_t = \mu(X_t, \theta)\,dt + \sigma(X_t, \theta)\,dW_t \tag{1}$$

that is, a joint parametric family:

$$\mathcal{P} \equiv \left\{ \left(\mu\left(\cdot, \theta\right), \sigma^2\left(\cdot, \theta\right) \right) \mid \theta \in \Theta \right\} \tag{2}$$

where Θ is a compact subset of $\mathbb{R}^K$.

If we believe that the true process is a diffusion with drift and diffusion functions $\left(\mu_0\left(\cdot\right), \sigma_0^2\left(\cdot\right)\right)$, a specification test asks whether there are values of the parameters in Θ for which the parametric model $\mathcal{P}$ is an acceptable representation of the true process, i.e., do the functions $\left(\mu_0\left(\cdot\right), \sigma_0^2\left(\cdot\right)\right)$ belong to the parametric family $\mathcal{P}$? Formally, the null and alternative hypotheses are:

$$\begin{aligned} H_0 &: \exists\, \theta_0 \in \Theta \quad \text{such that } \mu\left(\cdot, \theta_0\right) = \mu_0\left(\cdot\right) \text{ and } \sigma^2\left(\cdot, \theta_0\right) \\ &= \sigma_0^2\left(\cdot\right) \end{aligned} \tag{3}$$

$$H_1 : \left(\mu_0\left(\cdot\right), \sigma_0^2\left(\cdot\right)\right) \notin \mathcal{P}.$$

Because $\left(\mu_0\left(\cdot\right), \sigma_0^2\left(\cdot\right)\right)$ cannot easily be estimated directly from discretely sampled data, Aït-Sahalia (1996) proposed to test the parametric specification (2) using an indirect approach. Let $\pi(\cdot, \theta)$ denote the marginal density implied by the parametric model (1), and $p(\Delta, \cdot|\cdot, \theta)$ the transition density. Under regularity assumptions, $\left(\mu\left(\cdot, \theta\right), \sigma^2\left(\cdot, \theta\right)\right)$ will uniquely characterize the marginal and transition densities over discrete time intervals. For example, the Ornstein–Uhlenbeck process $dX_t = \beta\left(\alpha - X_t\right)dt + \gamma dW_t$ specified by Vasicek (1977) generates Gaussian marginal and transitional densities. The square-root process $dX_t = \beta\left(\alpha - X_t\right)dt + \gamma X_t^{1/2} dW_t$ used by Cox et al. (1985) yields a Gamma marginal and non-central chi-squared transitional densities.

More generally, any parametrization $\mathcal{P}$ of μ and σ^2 corresponds to a parametrization of the marginal and transitional densities:

$$\begin{aligned} \Pi \equiv &\left\{ \left(\pi\left(\cdot, \theta\right), p\left(\Delta, \cdot, |\cdot, \theta\right)\right) \mid \left(\mu\left(\cdot, \theta\right), \sigma^2\left(\cdot, \theta\right)\right) \right. \\ &\left. \in \mathcal{P}, \theta \in \Theta \right\}. \end{aligned} \tag{4}$$

While the direct estimation of μ and σ^2 with discrete data is problematic, the estimation of the densities explicitly take into account the discreteness of the data. The basic idea of Aït-Sahalia (1996) is to use the mapping between the drift and diffusion on the one hand, and the marginal and transitional densities on the other, to test the model's specification using densities at the observed discrete frequency (π, p) instead of the infinitesimal characteristics of the process (μ, σ^2). Aït-Sahalia (1996) proposed two tests, one based on the marginal density π, the other on the transition density p and derived their asymptotic distributions under the assumption that the process is stationary.

The marginal density test is based on comparing a nonparametric, kernel, estimator of the stationary or marginal density $\pi(\cdot)$ of

☆ Prepared for the Conference in Honor of Peter C.B. Phillips' 60th Birthday, held at Singapore Management University in July 2008. This research was partly funded by the NSF under grant DMS-0532370.

* Corresponding author. Tel.: +1 609 258 4015.
E-mail addresses: yacine@princeton.edu (Y. Aït-Sahalia), joon@indiana.edu (J.Y. Park).
1 Tel.: +1 812 856 0268.

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2012.01.030

the process to a parametric estimator of $\pi(\cdot, \theta)$ derived under the null hypothesis that the parametric model is correct. If the null is true, both estimators are consistent for the true marginal density and the test statistic, which is a distance measure between these two estimates, will be small. If the parametric model is incorrect, however, only the nonparametric estimator will be consistent and the test statistic will be large. Throughout the analysis, the maintained hypothesis is that the data series is stationary, so that notions such as a marginal stationary density are meaningful.

Following that work, Pritsker (1998) noted that, in small samples, the near-nonstationarity of empirical interest rate data can lead to an over-rejection of the null hypothesis when critical values are computed based on the asymptotic distribution derived under stationarity (see also Chapman and Pearson (2000), and Chen and Gao (2004b) for a critique of the critique). Corradi and Swanson (2005) proposed the use of the bootstrap in the context of that test. Other important contributions to the literature on testing the specification of diffusions include Hong and Li (2005), who use the fact that under the null hypothesis, the random variables $\{P(X_i|X_{i-\Delta}, \theta)\}$ are a sequence of i.i.d. uniform random variables, Fan and Zhang (2003), Chen and Gao (2004a), and Li and Tkacz (2006).

In this paper, we examine what happens to the test above in the misspecified situation where the diffusion process is, in fact, non-stationary, or nearly non-stationary. The latter situation is a good approximation for US interest rate time series. Of particular interest is therefore the question of whether near non-stationarity of the data can account for the over-rejection of the null hypothesis that has been documented in small samples. Our purpose is not to develop a new test; rather, we analyze the behavior of the existing marginal density-based test when we get at or near the end of its domain of application.

The types of non-stationarity we consider below relate naturally to similar concepts in classical discrete time series analysis such as unit root type behavior, or explosive behavior. The near non-stationarity we consider is essentially a continuous-time version of the "local to unity" analysis in the nearly integrated time series situation, which will be no stranger to readers familiar with P.C.B. Phillips' work. We focus solely here on the marginal density-based test because transition-based tests can be made to operate on increments of the process and as such do not rely as much on stationarity. (In fact, they can be designed to work perfectly with unit root data since increments of the process then become i.i.d.) For that reason, the nonstationarity or near-nonstationarity problem we are addressing is specific to the marginal density.

The rest of the paper is organized as follows. In order to analyze the behavior of the marginal-density test statistic, we will need to examine what becomes of the kernel density estimator from the diffusion's discrete sample when the diffusion is no longer stationary. Since there is no longer a stationary density, what does the kernel estimator of the marginal density converge to? Its limit involves the local time of the diffusion process and we start in Section 2 by reviewing its relevant properties. Section 3 derives the asymptotic behavior of the kernel density estimator. We can then apply these results to study the consistency of the test when the diffusion is non-stationary in Section 4, and its distribution when the diffusion is nearly integrated in Section 5. Using a setup that closely mimics the simulation design used by Pritsker (1998), who simulated Ornstein–Uhlenbeck processes with small values of mean reversion and diffusion parameters, we show that our asymptotic theory can explain why the marginal density-based test over-rejects in his simulations. Section 6 concludes. Proofs are in the Appendix.

2. Local time preliminaries

We assume that the data is a discrete sample from a time-homogeneous diffusion process X, solution of the stochastic differential equation

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t \tag{5}$$

where μ and σ^2 are, respectively, the drift and diffusion functions. We let $D = (\underline{x}, \bar{x})$ denote the domain of the diffusion X_t. In general, $D = (-\infty, +\infty)$, but in many examples in finance, X_t is the price of an asset with limited liability (stock, foreign currency, bond, etc.) or a nominal interest rate, in which case $D = (0, +\infty)$. We first assume that:

Assumption 1 (*Existence*)**.** The stochastic differential equation (5) admits a weak solution which is unique in probability law. The solution process $\{X_t, t \geq 0\}$ admits a transition density $p(t, y|x)$ which is continuously differentiable in t, infinitely differentiable in y and x on D.

Standard sufficient conditions are available (see e.g., Karatzas and Shreve (1991)) to guarantee that Assumption 1 is satisfied. Our subsequent theory relies on evaluating how much time a diffusion process spends near each possible value. This necessitates the use of the following notions.

2.1. Hitting times, boundary behavior, recurrence and transience

For the sake of completeness, we start by reviewing some relevant concepts pertaining to diffusions (and in some cases more general Markov processes) which we will use extensively in the rest of the paper. For more details, we refer to standard treatments such as Ethier and Kurtz (1986).

Define the hitting time of a point y in D as

$$T_y = \begin{cases} \infty & \text{if } X_t \neq y \text{ for all } t \geq 0 \\ \inf\{t \geq 0 \mid X_t = y\} & \text{otherwise.} \end{cases} \tag{6}$$

That is, T_y is the first time, if any, that the process reaches the value y.

Throughout, we assume that the process is regular in the interior of D, i.e.,

$$\Pr\left(T_y < \infty \mid X_0 = x\right) > 0 \tag{7}$$

for all x and y in the interior of D. Consider the scale function S whose derivative is given by

$$s(x) \equiv S'(x) = \exp\left(-2\int^x \frac{\mu(y)}{\sigma^2(y)}dy\right) \tag{8}$$

and the speed density

$$m(x) \equiv \frac{1}{(\sigma^2 s)(x)} \tag{9}$$

on D.

For $\underline{x} < a < b < \bar{x}$, the probability of reaching b before a, when starting from x in $[a, b]$ is

$$\Pr(T_b < T_a \mid X_0 = x) = \frac{S(x) - S(a)}{S(b) - S(a)}.$$

The boundary $\underline{x}$ (resp. $\bar{x}$) is unattainable (in finite expected time) if $\Pr\left(T_{\underline{x}} = \infty\right) = 1$ (resp. $\Pr(T_{\bar{x}} = \infty) = 1$).

From Feller's test for explosions (see e.g., Section 15.6 in Karlin and Taylor (1981)), $\underline{x}$ is unattainable when $\Sigma(\underline{x}) = \infty$, where

$$\begin{aligned}\Sigma(\underline{x}) &\equiv \int_{\underline{x}}^{x}\left\{\int_{v}^{x} m(u)\,du\right\} s(v)\,dv \\ &= \int_{\underline{x}}^{x}\left\{\int_{\underline{x}}^{u} s(v)\,dv\right\} m(u)\,du.\end{aligned}$$

By contrast, $\underline{x}$ is attainable in finite expected time if $\Sigma(\underline{x}) < \infty$. Similarly, the boundary $\bar{x}$ is unattainable if and only if $\Sigma(\bar{x}) = \infty$, where

$$\Sigma(\bar{x}) \equiv \int_x^{\bar{x}} \left\{ \int_x^v m(u)du \right\} s(v)\,dv = \int_x^{\bar{x}} \left\{ \int_u^{\bar{x}} s(v)dv \right\} m(u)\,du.$$

Among unattainable boundaries, natural boundaries can neither be reached in finite time nor can the diffusion be started there. Entrance boundaries such as 0^+, cannot be reached starting from an interior point in $D = (0, +\infty)$, but it is possible for X to start there in which case the process moves quickly away from 0 and never returns. Assuming $\bar{x}$ is unattainable, it is natural if $N(\bar{x}) = \infty$ and entrance if $N(\bar{x}) < \infty$ (see e.g., Table 6.2 in Karlin and Taylor (1981)), where

$$N(\bar{x}) \equiv \int_x^{\bar{x}} \left\{ \int_x^v s(u)\,du \right\} m(v)\,dv = \int_x^{\bar{x}} \left\{ \int_u^{\bar{x}} m(v)dv \right\} s(u)\,du.$$

Similarly, whether $\underline{x}$ is an entrance or a natural boundary depends upon whether $N(\underline{x}) < \infty$ or $N(\underline{x}) = \infty$ respectively, where

$$N(\underline{x}) \equiv \int_{\underline{x}}^x \left\{ \int_v^x s(u)\,du \right\} m(v)\,dv = \int_{\underline{x}}^x \left\{ \int_{\underline{x}}^u m(v)\,dv \right\} s(u)\,du.$$

A diffusion is *recurrent* iff

$$\Pr\left(T_y < \infty \mid X_0 = x\right) = 1 \tag{10}$$

for all x and y in the interior of D. A simple necessary and sufficient criterion for recurrence is as follows. Define the Green function

$$G_\alpha(y|x) = \int_0^{+\infty} e^{-\alpha t} p(y|x,t)\,dt. \tag{11}$$

A recurrent diffusion satisfies

$$\lim_{\alpha \searrow 0} G_\alpha(y|x) = \infty. \tag{12}$$

A diffusion which is not recurrent is said to be *transient*. Such a diffusion satisfies

$$0 < \Pr\left(T_y < \infty \mid X_0 = x\right) < 1 \tag{13}$$

for at least some x and y in the interior of D. A necessary and sufficient criterion for transience is

$$G_0(y|x) < \infty. \tag{14}$$

A recurrent diffusion is said to be *null recurrent* if $E\left[T_y \mid X_0 = x\right] = \infty$ for all x and y in the interior of D, and *positively recurrent* if $E\left[T_y \mid X_0 = x\right] < \infty$ for all x and y in the interior of D. A recurrent diffusion is positively recurrent iff the speed measure m is integrable in x at both boundaries of D i.e.,

$$\int_D m(y)dy < \infty \tag{15}$$

and then

$$\lim_{\alpha \searrow 0} \alpha G_\alpha(y|x) = \frac{1}{\int_D m(y)dy}. \tag{16}$$

When (15) is satisfied, let

$$\pi(x) \equiv \frac{m(x)}{\int_D m(y)dy}. \tag{17}$$

If the initial value of the process, X_0, has density π, then the process is stationary and $\pi(x)$ is its stationary density, i.e., the common marginal density of each X_t.

Stationary diffusions are recurrent (with the proper initialization and support definition), but the converse is not necessarily true. When both boundaries of the process are entrance boundaries, (15) is automatically satisfied. When at least one of the boundaries is natural, stationarity is neither precluded nor implied in that the (only) possible candidate for stationary density, π, or equivalently m, may or may not be integrable near the boundaries. For instance, both a mean-reverting Ornstein–Uhlenbeck process,

$$dX_t = -\beta X_t dt + \gamma dW_t \tag{18}$$

with $\beta > 0$, and a Brownian motion have natural boundaries at $-\infty$ and $+\infty$. Yet the former process is positively recurrent, due to mean-reversion, while the latter is null recurrent and not stationary.

2.2. The local time of a diffusion process

As will become apparent below, our analysis makes heavy uses of the concept of local time. This is of course a classical concept in the analysis of stochastic processes. For its use in situations related to the present paper, see, e.g., Akonom (1993) and Bandi and Phillips (2009). The local time L of X at an interior point x of D is defined as

$$L(T,x) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^T 1\{|X_t - x| < \varepsilon\}\,dt \tag{19}$$

(see e.g., Section 15.12.D of Karlin and Taylor (1981)). Intuitively, $L(T,x)$ denote the time spent by X in the neighborhood of point x between time 0 and time T. For this reason, it is often called the sojourn time. It is more common in the stochastic process literature to define the local time to be the sojourn time measured by the quadratic variation of the underlying process. However, our definition is more appropriate for the theory developed in the paper.

It is well known that the local time L exists with a version continuous a.s. in both arguments for continuous semimartingales. The local time L satisfies Tanaka's occupation times formula

$$\int_0^T f(X_t)dt = \int_{-\infty}^{\infty} f(x)L(T,x)dx \tag{20}$$

for any nonnegative function f on $\mathbb{R}$, which allows us to switch between time and space integrals.

The local time L is a random variable whose first two moments are given by:

Lemma 1. *The conditional expected value and variance of the local time of the process are related to its transition density through*

$$E\left[L(T,y) \mid X_0 = x\right] = \int_0^T p(t,y|x)dt \tag{21}$$

$$E\left[L(T,y)^2 \mid X_0 = x\right] = 2\int_0^T \int_0^t p(s,y|x)p(t-s,y|y)dsdt. \tag{22}$$

The expected local time and its variants for general semimartingales are analyzed extensively in Park (2007), which shows in particular how we may estimate them nonparametrically from discretely observed samples and do inferences on them using their nonparametric estimates.

Throughout the paper, we use the notation "$=_d$" to denote equality in distribution and "$\approx$" to indicate *equality of the leading terms*, i.e., $P \approx Q$ ($P \approx Q$ a.s.) means that $P = Q(1 + o_p(1))$ ($P = Q(1+o(1))$ a.s.), and "$\sim$" to indicate *equality of the stochastic orders*, i.e., $P \sim Q$ implies $P/Q = Q/P = O_p(1)$.

2.3. Asymptotic behavior of the local time over the sampling horizon

Of critical importance in this paper will be the behavior of the local time of the diffusion asymptotically in T. As we shall see, it controls the asymptotic behavior of the test statistic under consideration. We make the following assumption, and then show that it is satisfied for many processes of interest:

Assumption 2. There exists $\kappa \in [0, 1]$ such that for each x in D, $L(T, x) \sim T^{\kappa}$ as $T \to \infty$.

The condition in Assumption 2 holds for all diffusions satisfying the Darling–Kac condition; see Darling and Kac (1957) and the discussions in Bandi and Moloche (2005) and Bandi and Phillips (2009). In general, we have $L(T, x) \sim T^{\kappa} l(T)$ for $0 \le \kappa \le 1$, where l is varying slowly at infinity, i.e., $l(\lambda T)/l(T) \to 1$ as $T \to \infty$ for all $\lambda > 0$. However, we ignore the slowly varying component to simplify the presentation of the results in the paper. The reader is referred to Höpfner and Löcherbach (2003) and the references cited there for more details on the limit theory of local times for general Markov processes.

As mentioned, $L(T, x)$ measures how much time the process spends near x between times 0 and T. For instance, a stationary process should be expected to spend more time near any given value x than an explosive process, which barely visits each point. So the time order κ of the local time should be related to the recurrence property of the underlying diffusion X. Indeed, we will see that $\kappa = 1$ for stationary or positive recurrent diffusions, and $\kappa = 0$ for transient processes. We will also give examples illustrating the intermediary range $0 < \kappa < 1$, including Brownian motion for which $\kappa = 1/2$.

In the case of stationary diffusions, we have as $T \to \infty$

$$\frac{L(T, x)}{T} \to \pi(x) \quad \text{a.s.} \tag{23}$$

as shown in, e.g., Theorem 6.3 of Bosq (1998). Assumption 2 is thus satisfied with $\kappa = 1$ for every $x \in D$.

As discussed, Brownian motion is null recurrent (and $G_{\alpha}(y|x) = \exp(-\sqrt{2\alpha}\,|y - x|)/\sqrt{2\alpha}$). For this process, the assumption holds with $\kappa = 1/2$, since by invariance scaling we have in this case

$$L(T, x) =_d T^{1/2} L(1, T^{-1/2} x) \tag{24}$$

(see e.g., Exercise VI.2.11 in Revuz and Yor (1994)), and for fixed x, as $T \to \infty$, we have

$$L(1, T^{-1/2} x) \to L(1, 0) \quad \text{a.s.,}$$

due to the continuity of $L(1, \cdot)$. The distribution of $L(1, 0)$ or more generally $L(T, 0)$ is given by (see e.g., Exercise VI.2.18 in Revuz and Yor (1994))

$$\Pr\left(L(T, 0) \ge u\right) = \sqrt{\frac{2}{\pi T}} \int_u^{+\infty} \exp\left(-\frac{y^2}{2T}\right) dy$$

for $u \ge 0$ (Paul Lévy showed that $L(T, 0)$ and $\sup_{0 \le t \le T} W_t$ have the same distribution.)

As an example of other nonstationary processes, we consider martingale diffusions

$$dX_t = c|X_t|^{-r/2} dW_t \tag{25}$$

with $r > -1$ and some constant $c > 0$. They yield a class of processes with κ ranging from 0 to 1 as shown below:

Lemma 2. *If X is generated as in (25), then we have*

$$L(T, x) =_d T^{\kappa} L(1, T^{-(1-\kappa)} x)$$

for all $x \in D$, where $\kappa = (r + 1)/(r + 2)$ with $r > -1$.

For the martingale diffusion given by (25), κ ranges from 0 to 1 as r increases from -1 to infinity. Of course, Brownian motion can be considered as a special case with $\kappa = 1/2$, which yields (24). As in the case of Brownian motion, $L(1, \cdot)$ is continuous and we have

$$L(1, T^{-(1-\kappa)} x) \to L(1, 0) \quad \text{a.s.,} \tag{26}$$

as $T \to \infty$.

In the transient case, we have $\kappa = 0$ and from (21) and (14):

$$E\left[L(\infty, y) \mid X_0 = x\right] = G_0(y|x).$$

For example, for Brownian motion with drift, $dX_t = \mu dt + dW_t$, $\mu > 0$, we have

$$G_0(y|x) = \begin{cases} 1/\mu & \text{if } y \ge x \\ \exp(2\mu(y - x))/\mu & \text{if } y \le x \end{cases} \tag{27}$$

and furthermore

$$\begin{aligned} &\Pr\left(L(\infty, y) > u \mid X_0 = x\right) \\ &\quad = \begin{cases} \exp(-\mu u) & \text{if } y \ge x \\ \exp(-\mu \exp(-2\mu(y - x))u) & \text{if } y \le x. \end{cases} \end{aligned} \tag{28}$$

Other examples of transient processes include the Ornstein–Uhlenbeck process with negative coefficient of mean reversion ($\beta < 0$ in (18)), and Bessel processes

$$dX_t = \frac{\delta - 1}{2X_t} dt + dW_t \tag{29}$$

with dimension parameter $\delta > 2$. Next, we assume:

Assumption 3. Let κ be defined as in Assumption 2. Moreover, let a_p be homogeneous of degree p for $p = 0, 1, 2$, and let b be a bounded integrable function. Assume for $q = 1, 2, 3, 4$

$$\int_{-\infty}^{\infty} a_p(x) L^q(T, x)\, dx \sim T^{(p+1)(1-\kappa)+q\kappa} \tag{30}$$

$$\int_{-\infty}^{\infty} b(x) L^q(T, x)\, dx \sim T^{q\kappa} \tag{31}$$

as $T \to \infty$.

It is possible to establish (instead of assuming) Assumption 3 rigorously from first principles, at least in some special cases, and we will do so below. But attempting this in general does not seem desirable, since our main focus is not to establish this result but rather to use it. Roughly speaking, Assumption 3 requires that the divergence rate of $L(T, x)$ as $T \to \infty$ be the same for all x's. In this case, we may write

$$L(T, x) = T^{\kappa} \ell_T(x). \tag{32}$$

But we must also have

$$\int_{-\infty}^{\infty} L(T, x) dx \equiv T,$$

since the total time spent between 0 and T in the neighborhood of *all* points in D is T (this is of course consistent with applying (20) to $f \equiv 1$), or equivalently

$$\int_{-\infty}^{\infty} \ell_T(x) dx \equiv T^{1-\kappa}, \tag{33}$$

where ℓ_T is neither exploding nor vanishing on its support in light of Assumption 2 and (32); for that to happen together with (33), we expect the support of ℓ_T to expand at rate $T^{1-\kappa}$ as $T \to \infty$. On its support, $\ell_T^q(x) \sim 1$ and we then have

$$\int_{-T^{1-\kappa}}^{T^{1-\kappa}} x^p \ell_T^q(x) dx \sim \int_{-T^{1-\kappa}}^{T^{1-\kappa}} x^p dx \sim T^{(p+1)(1-\kappa)}.$$

We may therefore expect to have

$$\int_{-\infty}^{\infty} a_p(x)L^q(T,x)\,dx \sim T^{q\kappa}\int_{-T^{1-\kappa}}^{T^{1-\kappa}} x^p \ell_T^q(x)dx \sim T^{(p+1)(1-\kappa)+q\kappa}$$

and similarly

$$\int_{-\infty}^{\infty} b(x)L^q(T,x)\,dx \sim T^{q\kappa}\int_{-\infty}^{\infty} b(x)\ell_T^q(x)dx \sim T^{q\kappa}$$

as we assume in Assumption 3.

We can verify that Assumption 3 is satisfied using a direct calculation in some cases. For a stationary diffusion, we have $\kappa = 1$ and (30)–(31) are satisfied as long as the diffusion has a stationary distribution π that has bounded density and finite variance. Indeed, if we write for $f = a_p$ or b, then from (23) we have

$$\int_{-\infty}^{\infty} f(x)L^q(T,x)\,dx = T^{q\kappa}\int_{-\infty}^{\infty} f(x)\left[\frac{L(T,x)}{T}\right]^q dx \sim T^{q\kappa}\int_{-\infty}^{\infty} f(x)\pi^q(x)dx$$

and $\lim \pi(x) = 0$ as $x \to \infty$ implies convergence of the integral above for all values of $q \geq 1$ provided that the density π admits a finite variance (since $p \leq 2$).

For the martingale diffusion given by (25), we have $0 < \kappa < 1$ and it follows from Lemma 2 that

$$\int_{-\infty}^{\infty} a_p(x)L^q(T,x)\,dx =_d T^{q\kappa}\int_{-\infty}^{\infty} a_p(x)L^q(1,T^{-(1-\kappa)}x)dx = T^{(p+1)(1-\kappa)+q\kappa}\int_{-\infty}^{\infty} a_p(y)L^q(1,y)dy \quad (34)$$

using the change of variable $y = T^{-(1-\kappa)}x$. Similarly, we have from Lemma 2 and (26)

$$\int_{-\infty}^{\infty} b(x)L^q(T,x)\,dx =_d T^{q\kappa}\int_{-\infty}^{\infty} b(x)L^q(1,T^{-(1-\kappa)}x)\,dx \approx T^{q\kappa}L^q(1,0)\int_{-\infty}^{\infty} b(x)\,dx \quad \text{a.s.} \quad (35)$$

by dominated convergence. The conditions in Assumption 3 are thus satisfied by the martingale diffusion including, of course, Brownian motion with $\kappa = 1/2$.

For Brownian motion with drift, we have $\kappa = 0$ and we can compute directly

$$E\left[\int_{-\infty}^{\infty} a_p(y)L^q(T,y)\,dy\,\middle|\, X_0 = x\right] = \int_{-\infty}^{\infty} a_p(y)E\left[L^q(T,y)\mid X_0 = y\right]dy$$

and use Lemma 1; for instance when $q = 1$, we have

$$\begin{aligned}\int_{-\infty}^{\infty} y^pE\left[L(T,y)\mid X_0 = x\right]dy &= \int_{-\infty}^{\infty} y^p\int_0^T p(t,y|x)dtdy \\ &= \int_0^T\int_{-\infty}^{\infty} y^p p(t,y|x)dydt \\ &= \begin{cases} T & \text{if } p = 0 \\ xT + \dfrac{\mu}{2}T^2 & \text{if } p = 1 \\ x^2T + \dfrac{(1+2x\mu)}{3}T^2 + \dfrac{\mu^2}{3}T^3 & \text{if } p = 2\end{cases}\end{aligned}$$

since $p(t,y|x)$ is the Gaussian density with mean $x + \mu t$ and variance t. When $q = 2$,

$$\begin{aligned}\int_{-\infty}^{\infty} y^pE\left[L^2(T,y)\mid X_0 = x\right]dy &= 2\int_{-\infty}^{\infty} y^p\int_0^T\int_0^t p(s,y|x)p(t-s,y|y)dsdtdy \\ &= 2\int_0^T\int_0^t\int_{-\infty}^{\infty} y^p p(s,y|x)p(t-s,y|y)dydsdt \\ &= \begin{cases} \dfrac{2\,\mathrm{erf}\left(\frac{\sqrt{T}\mu}{\sqrt{2}}\right)}{\mu}T + O(1) & \text{if } p = 0 \\ \mathrm{erf}\left(\dfrac{\sqrt{T}\mu}{\sqrt{2}}\right)T^2 + O(T) & \text{if } p = 1 \\ \dfrac{2\mu\,\mathrm{erf}\left(\frac{\sqrt{T}\mu}{\sqrt{2}}\right)}{3}T^3 + O\left(T^2\right) & \text{if } p = 2\end{cases}\end{aligned}$$

where erf designates the normal error function, which tends to 1 exponentially fast as T tends to infinity. The perhaps more familiar normal c.d.f. N is related to erf by

$$N(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right).$$

The verification for $q > 2$ proceeds similarly. Since by Chebyshev's Inequality, $E\left[|A_T|\right] = O(a_T)$ implies that $A_T \sim a_T$, the conditions in Assumption 3 follow for Brownian motion with drift.

3. Asymptotics for the kernel density and local time estimators

We now have the tools we need to understand what is estimated by the kernel marginal density estimator that enters the test statistic. Recall that the process X is assumed to be observed at intervals of length Δ over time $[0,T]$. The sample size n is, therefore, given by $n = T/\Delta$. Here we assume that the observation intervals are equispaced. However, this is just for expositional simplicity. All of our subsequent results can easily be extended to more general cases where X are observed irregularly, i.e., at $\Delta_1, \Delta_1 + \Delta_2, \ldots$ and so on, as long as the maximum of the observation intervals, $\max_{1\leq i\leq n}\Delta_i$, decreases down to zero. For the development of our asymptotics, we let $T \to \infty$ and $\Delta \to 0$ so that $n \to \infty$. The reliance on asymptotics where $\Delta \to 0$ is dictated by the need to allow for stationary as well as nonstationary dynamics; this is consistent with earlier approaches as discussed in the review paper by Bandi and Phillips (2009).

We first consider the asymptotics of the sample moments $\sum_{i=1}^n f(X_{i\Delta})$ for a real valued function f on $\mathbb{R}$. The following result is fairly general and may potentially be useful beyond the specific scope of its application here to the test statistic of interest:

Lemma 3. *Let $f = a_p$ or b, where a_p and b are defined in Assumption 3. Suppose that f is twice continuously differentiable, and assume either*

(a) $f = a_p$ and $|\mu(x)| \leq c_1|x|^r$, $|\sigma^2(x)| \leq c_2|x|^{r+1}$ for some constants c_1 and c_2 with r satisfying $\Delta T^{(r-1)(1-\kappa)} \to 0$, or

(b) $f = b$ has derivatives b' and b'' such that $\mu b'$, $\sigma b'$ and $\sigma^2 b''$ are bounded and integrable.

Then we have

$$\Delta\sum_{i=1}^n f(X_{i\Delta}) \approx \int_0^T f(X_t)\,dt \quad \text{a.s.} \quad (36)$$

uniformly in T.

For a wide class of f, Lemma 3 provides the precise conditions sufficient to have the continuous approximation in (36) for the sample moments given by the discrete observations of the underlying diffusions. Of course, here we require $\Delta \to 0$, i.e., the infills should take place to have the continuous approximation. The required rates of infills are, however, different depending upon the class of functions f and the explosiveness of the drift and diffusion functions. If f is a homogeneous function and if the drift function μ or the diffusion function σ is more explosive than the linear specification with $r > 1$, then Δ must decrease faster than $T^{-(r-1)(1-\kappa)}$. The required rate increases, as the drift or diffusion function becomes more explosive and/or the underlying diffusion becomes more transient. On the other hand, required is no extra condition other than $\Delta \to 0$ if μ and σ are bounded by a linear function, and/or the underlying diffusion is stationary with $\kappa = 1$. For integrable and bounded f, $\Delta \to 0$ alone is sufficient for the continuous approximation if it has derivatives vanishing fast enough at infinity, compared to the rates at which the drift and diffusion functions explode.

In light of the occupation times formula (20), the stochastic order of the continuous sample moment in (36) can now be obtained directly from Assumption 3. Indeed, in some special cases, we may deduce more explicit asymptotics using the local time asymptotics in (23), (26) and Lemma 2. For stationary ergodic diffusions, we have

$$\begin{aligned}\frac{\Delta}{T}\sum_{i=1}^{n} f(X_{i\Delta}) &= \frac{1}{T}\int_0^T f(X_t)\,dt + o(1) \quad \text{a.s.}\\ &= \int_{-\infty}^{\infty} f(x)\frac{L(T,x)}{T}\,dx + o(1) \quad \text{a.s.}\\ &\to \int_{-\infty}^{\infty} f(x)\pi(x)\,dx \quad \text{a.s.}\end{aligned}$$

for any f if the underlying density π admits the finite expectation of f. Likewise, for the martingale diffusion (25), we have for $f = a_p$

$$\begin{aligned}&\frac{\Delta}{T^{p(1-\kappa)+1}}\sum_{i=1}^{n} a_p(X_{i\Delta})\\ &= \frac{1}{T^{p(1-\kappa)+1}}\int_0^T a_p(X_t)\,dt + o(1) \quad \text{a.s.}\\ &= \frac{1}{T^{p(1-\kappa)+1}}\int_{-\infty}^{\infty} a_p(x)L(T,x)\,dx + o(1) \quad \text{a.s.}\\ &\to_d \int_{-\infty}^{\infty} a_p(x)L(1,x)\,dx\end{aligned}$$

and for $f = b$

$$\begin{aligned}\frac{\Delta}{T^{\kappa}}\sum_{i=1}^{n} b(X_{i\Delta}) &= \frac{1}{T^{\kappa}}\int_0^T b(X_t)\,dt + o(1) \quad \text{a.s.}\\ &= \frac{1}{T^{\kappa}}\int_{-\infty}^{\infty} b(x)L(T,x)\,dx + o(1) \quad \text{a.s.}\\ &\to_d L(1,0)\int_{-\infty}^{\infty} b(x)\,dx\end{aligned}$$

due, respectively, to the results in (34) and (35).

We now consider the asymptotics for the kernel density estimator. Strictly speaking, the kernel density estimator would make sense only for stationary processes, since the time invariant marginal density does not exist for nonstationary processes. But we may look more generally at the kernel estimator for the local time $L(T,x)$, which is given by

$$\hat{L}(T,x) = \frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right), \tag{37}$$

where K is the kernel function and h is the bandwidth parameter. As we show below, the local time estimator in (37) provides a consistent estimator of local time for both stationary and nonstationary processes.

The kernel estimator for the local time is indeed nothing but the T-multiple of the usual kernel density estimator

$$\hat{\pi}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right) \tag{38}$$

which is commonly used to estimate the time invariant marginal density under stationarity, and constitutes the nonparametric estimator in the test statistic whose behavior we would like to investigate. Under stationarity and ergodicity, we may deduce in particular that

$$\hat{\pi}(x) = \frac{\hat{L}(T,x)}{T} = \frac{L(T,x)}{T} + o(1)\text{a.s.} \to \pi(x) \quad \text{a.s.}$$

from Lemma 4 and the local time asymptotics in (23) introduced earlier.

For the kernel function K, we impose the usual conditions.

Assumption 4. We assume that K is bounded and twice continuously differentiable, and that $\int_{-\infty}^{\infty} K(x)\,dx = 1$, $\int_{-\infty}^{\infty} xK(x)\,dx = 0$, and $\int_{-\infty}^{\infty} x^2|K(x)|\,dx < \infty$.

For the local time estimator $\hat{L}(T,x)$ in (37), we may well expect to have continuous approximation similar to (36), if we choose h appropriately as a function of Δ and T. Once we establish the required continuous approximation, we may easily deduce the consistency of $\hat{L}(T,x)$, since

$$\begin{aligned}\hat{L}(T,x) &\approx \frac{1}{h}\int_0^T K\left(\frac{X_t-x}{h}\right)dt\\ &= \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{u-x}{h}\right)L(T,u)\,du\\ &= \int_{-\infty}^{\infty} K(u)L(T,x+hu)\,du \approx L(T,x)\end{aligned}$$

by successively applying occupation times formula and change of variable, and using the continuity of $L(T,\cdot)$.

Lemma 4. *Let h be chosen such that $h \to 0$ and $\Delta/h^2 \to 0$ as $\Delta \to 0$. Then we have*

$$\hat{L}(T,x) \approx L(T,x) \quad \text{a.s.}$$

for every T and x. Moreover, we let $f = a_p$ or b with a_p and b defined in Assumption 3, and define $m = [(\mu/\sigma^2)' + (\mu/\sigma^2)^2][(1/\sigma^2) + (1/\sigma^2)'']$. Suppose that we set $hL(T,x)^{1/3-\delta} \to \infty$ a.s. for some $\delta > 0$ and assume either

(a) $f = a_p$ and $|m(x)| \le c_1|x|^{r_1}$, $|\mu(x)| \le c_2|x|^{r_2}$ and $|\sigma^2(x)| \le c_3|x|^{r_3}$ for some constants c_1, c_2 and c_3 with r_1, r_2 and r_3 satisfying $h^2T^{r_1(1-\kappa)} \to 0$, $\Delta h^{-1}T^{r_2(1-\kappa)} \to 0$, $\Delta h^{-2}T^{r_3(1-\kappa)} \to 0$ for $h \to 0$, or

(b) $f = b$ with mb, μb and $\sigma^2 b$ bounded and integrable, and $h \to 0$ such that $\Delta/h^2 \to 0$.

Then we have

$$\int_{-\infty}^{\infty} f(x)\hat{L}^q(T,x)\,dx \approx \int_{-\infty}^{\infty} f(x)L^q(T,x)\,dx \quad \text{a.s.}$$

uniformly in T.

The consistency of the kernel estimator for local time in (37) was established earlier in Bandi and Phillips (2003) under comparable conditions, and later by Aït-Sahalia and Park (2009) under the same conditions imposed here. In this paper, we extend their results to the functionals that involve the powers of local time. Our result here can be used to analyze the statistics given, under the assumption of stationarity, as functionals of the estimated density. This will be clearly seen in the next section.

We may use the results in Lemma 4 to establish more explicit asymptotics in the special cases that we considered following Lemma 3. For instance, if we have stationary diffusions, then

$$\begin{aligned}
&\frac{1}{T^q}\int_{-\infty}^{\infty} f(x)\hat{L}^q(T,x)\,dx \\
&\quad= \frac{1}{T^q}\int_{-\infty}^{\infty} f(x)L^q(T,x)\,dx + o(1) \quad \text{a.s.} \\
&\quad= \int_{-\infty}^{\infty} f(x)\left[\frac{L(T,x)}{T}\right]^q dx + o(1) \quad \text{a.s.} \\
&\quad\to \int_{-\infty}^{\infty} f(x)\pi^q(x)\,dx \quad \text{a.s.}
\end{aligned}$$

if for instance π is bounded and f admits finite expectation. For the martingale diffusion (25), we have for $f = a_p$

$$\begin{aligned}
&\frac{1}{T^{(p+1)(1-\kappa)+q\kappa}}\int_{-\infty}^{\infty} a_p(x)\hat{L}^q(T,x)\,dx \\
&\quad= \frac{1}{T^{(p+1)(1-\kappa)+q\kappa}}\int_{-\infty}^{\infty} a_p(x)L^q(T,x)\,dx + o(1) \quad \text{a.s.} \\
&\quad\to_d \int_{-\infty}^{\infty} a_p(x)L^q(1,x)\,dx
\end{aligned}$$

and for $f = b$

$$\begin{aligned}
&\frac{1}{T^{q\kappa}}\int_{-\infty}^{\infty} b(x)\hat{L}^q(T,x)\,dx \\
&\quad= \frac{1}{T^{q\kappa}}\int_{-\infty}^{\infty} b(x)L^q(T,x)\,dx + o(1) \quad \text{a.s.} \\
&\quad\to_d L^q(1,0)\int_{-\infty}^{\infty} b(x)\,dx
\end{aligned}$$

using the results in (34) and (35).

4. Consistency of the marginal specification test when the diffusion is nonstationary

We now consider the marginal density-based specification test proposed by Aït-Sahalia (1996). Let

$$\Pi_M \equiv \left\{\pi(\cdot,\theta) \mid \left(\mu(\cdot,\theta), \sigma^2(\cdot,\theta)\right) \in \mathcal{P}, \theta \in \Theta\right\}$$

denote the parametric family of marginal densities implied by the specification of the parametric model (1). This family is characterized by the fact that the density $\pi(\cdot,\theta)$ corresponding to the pair $\left(\mu(\cdot,\theta), \sigma^2(\cdot,\theta)\right)$ is:

$$\pi(x,\theta) = \frac{\xi(\theta)}{\sigma^2(x,\theta)} \exp\left\{\int^x \frac{2\mu(u,\theta)}{\sigma^2(u,\theta)}du\right\} \tag{39}$$

where the choice of the lower bound of integration in the interior of the domain of the diffusion is irrelevant, and is absorbed in the normalization constant $\xi(\theta)$ determined to insure that the density integrates to one. If we let the true marginal density of the process be

$$\pi_0(x) = \frac{\xi_0}{\sigma_0^2(x)} \exp\left\{\int^x \frac{2\mu_0(u)}{\sigma_0^2(u)}du\right\} \tag{40}$$

we can then test

$$\begin{aligned}
&H_{M0}: \exists\, \theta_0 \in \Theta \quad \text{such that } \pi(\cdot,\theta_0) = \pi_0(\cdot) \\
&H_{M1}: \pi_0(\cdot) \notin \Pi_M.
\end{aligned} \tag{41}$$

It is necessary that H_{M0} be true for H_0 to be true. If the true density $\pi_0(\cdot)$ were known, we could simply check to see if it belonged to the proposed parametric class. Since it is unknown, we must estimate it, and do so with an estimator that does not already assume that the null hypothesis is correct (otherwise there is obviously no way of testing it). We use for that purpose a nonparametric estimator – that is, free of all parametric assumptions regarding μ and σ^2 – which will converge to the true density whether or not the parametric model (1) is correctly specified. Now consider a parametric estimator of the implied density $\pi(\cdot,\theta_0)$. It will converge to the true density only if the model is correctly specified. Therefore the parametric and nonparametric density estimators should be close together if the parametric model is correct, and far from each other otherwise. A measure of distance M between the two density estimates provides a natural statistic to test the null hypothesis of correct parametric specification. Aït-Sahalia (1996) suggested to test H_{M0} using the distance measure between the densities:

$$M \equiv \min_{\theta\in\Theta} \int_{\underline{x}}^{\bar{x}} \left(\pi(x,\theta) - \pi_0(x)\right)^2 \pi_0(x)\, w(x)dx \tag{42}$$

where w is a weight (or trimming) function. Here we consider $w(x) = 1\{|x| \le c\}$ for some constant c or $w(x) = 1$. The former is used to analyze the *fixed-point* trimming, and the second no trimming or the *fixed-percentage* trimming. (When D has finite boundaries, trimming can be advisable.)

The test statistic is based on

$$\hat{M} = nh \min_{\theta\in\Theta} \int_D \left(\pi(x,\theta) - \hat{\pi}(x)\right)^2 \hat{\pi}(x) w(x)dx, \tag{43}$$

where $\hat{\pi}$ is the nonparametric density estimate of π_0, computed with bandwidth h. The null hypothesis is therefore rejected when the test statistic $\hat{M}$ is large enough. The test evaluates the distance between the densities at the "best possible" parametric estimator:

$$\hat{\theta}_M \equiv \operatorname*{argmin}_{\theta\in\Theta} \int_D \left(\pi(x,\theta) - \hat{\pi}(x)\right)^2 \hat{\pi}(x) w(x)dx. \tag{44}$$

The test rejects the parametric specification $\pi(\cdot,\theta)$ if $\hat{M}$ takes a large value. The intuition for the test is straightforward. If $\pi(\cdot,\theta)$ is the correct specification, $\hat{M}$ should be small.

Under stationarity, $\hat{M}$ is shown in Aït-Sahalia (1996) to be asymptotically normally distributed and the critical value $c(\alpha)$ of the size α test is given by

$$\hat{c}(\alpha) = \hat{E}_M + z_{1-\alpha} h^{1/2} \hat{V}_M^{-1/2},$$

where $z_{1-\alpha}$ is the one-sided normal cutoff for α, and

$$\begin{aligned}
&\hat{E}_M = c_1 \int_{-\infty}^{\infty} \hat{\pi}^2(x) w(x)\,dx \quad \text{and} \\
&\hat{V}_M = c_2 \int_{-\infty}^{\infty} \hat{\pi}^4(x) w(x)\,dx
\end{aligned} \tag{45}$$

for constants c_1 and c_2 dependent upon the kernel function only, and given by

$$\begin{aligned}
&c_1 = \int_{-\infty}^{+\infty} K^2(x)\,dx \\
&c_2 = 2\int_{-\infty}^{+\infty} \left\{\int_{-\infty}^{+\infty} K(u)K(u+x)\,du\right\}^2 dx.
\end{aligned}$$

It was then suggested that h be chosen satisfying

$$h = c\,\hat{\sigma} n^{-2/9}/\ln(n), \tag{46}$$

where $\hat{\sigma}$ is the standard deviation of the data, and c is a constant.

In this paper, we investigate the test consistency under nonstationarity. For nonstationary diffusions, the time invariant marginal distribution does not exist. Naturally, any parametric family of probability densities cannot correctly specify the marginal distribution that changes over time. The test therefore must be able to reject the null of correct specification against any parametric family of probability densities in this case. Though our subsequent analysis will mainly be focused on nonstationary diffusions, we do not exclude stationary diffusions. In fact, we allow for stationary, as well as nonstationary, diffusions, and our subsequent results are applicable for both stationary and nonstationary diffusions. We first concentrate exclusively on the consistency of the test. Therefore, we will maintain that the models are misspecified also for stationary diffusions, i.e., the true marginal distribution is *not* represented by a member of the family of probability densities $\pi(\cdot, \theta)$ with some $\theta \in \Theta$.

Assumption 5. We assume that Θ is compact, $\pi(x, \theta)$ is continuous in $\theta \in \Theta$ for all $x \in D$, and $\sup_{\theta\in\Theta} \pi(\cdot, \theta)$ is bounded and integrable.

The conditions introduced in Assumption 5 are standard, and necessary to warrant the uniform convergence of the objective function in (43). In the following propositions, we let h, Δ and T satisfy the conditions introduced in Lemmas 3 and 4.

Proposition 1. *Let $w(x) \equiv 1$. Then the test becomes consistent if*

$$\Delta = o(hT + h^{1/2}T^{(5\kappa-3)/2}),$$

which in turn holds if, for some $\delta > 0$,

$$\Delta = o(T^{(9\kappa-5)/4-\delta}),$$

when h is selected according to (46) and $\Delta T^r \to \infty$ for some $r > 0$.

Proposition 2. *Let $w(x) = 1\{|x| \leq c\}$ for some constant c. Then the test becomes consistent if*

$$\Delta = o(hT^{2-\kappa} + h^{1/2}T^{3\kappa-2}),$$

which in turn holds if, for some $\delta > 0$,

$$\Delta = o(T^{(45\kappa-29)/16-\delta}),$$

when h is selected according to (46) and $\Delta T^r \to \infty$ for some $r > 0$.

In both cases, the test becomes consistent also for nonstationary diffusions, as long as Δ is sufficiently small relative to T. It is clear, however, from the results in Propositions 1 and 2 that we may expect more powers of the test as Δ gets smaller compared to T. This seems very natural especially for nonstationary diffusions. Note that the discrete samples provide more information on nonstationary diffusions as Δ decreases, while the underlying diffusions necessarily exhibit more nonstationary characteristics as T increases. Propositions 1 and 2 specify explicitly the relative magnitudes of T and Δ, for which we have enough information about the underlying nonstationary diffusions to reject any incorrect specifications of them as stationary diffusions. Of the tests relying on two different trimming schemes, we require more stringent conditions for the 'fixed-point' trimming than for no trimming or the 'fixed-percentage' trimming.

For stationary diffusions, both Propositions 1 and 2 imply that the misspecified models are rejected asymptotically with probability one under the minimal condition $T \to \infty$ or $\Delta \to 0$. No extra conditions are required for stationary diffusions. Indeed, we have in this case $\kappa = 1$, and the conditions required for the test consistency in both Propositions 1 and 2 reduce to $\Delta = o(T^{\delta})$ for some $\delta > 0$.

As κ decreases down from unity, we need more stringent conditions for the test consistency. This might be well expected since as κ becomes smaller the marginal density, which is just a scaled local time, generally becomes less informative on the marginal distribution of the underlying diffusion. For Brownian motions, we have $\kappa = 1/2$ and the required condition for the test consistency becomes

$$\Delta = o(T^{-1/8-\delta}) \quad \text{or} \quad \Delta = o(T^{-13/32-\delta})$$

depending upon whether which of the two weight functions is used. In most of practical applications, we have the value of the observation interval Δ much smaller than the reciprocal of any fractional power of the time span T. If we use daily observation, for instance, then $\Delta \approx 1/250 = 0.004$. However, even for one hundred years of time span T, we have $T^{-1/8} \approx 0.5623$, which is relatively much bigger. For the usual values of T and Δ, the test is likely to have the asymptotic power close to one against Brownian motions.

The condition becomes most stringent for the transient diffusions with $\kappa = 0$. In this case, we must have

$$\Delta = o(T^{-5/4-\delta}) \quad \text{or} \quad \Delta = o(T^{-29/16-\delta})$$

for each of the tests with two different weight functions.

These results can also be useful to predict the performance of the test for stationary diffusions which are nearly nonstationary. For instance, an Ornstein–Uhlenbeck process with small mean-reversion parameter would certainly behave similarly as a Brownian motion. Therefore, it is expected to reject the null against any parametric specification. This point is well illustrated in Pritsker (1998). He shows through simulations that the test rejects the correct specification too often when the underlying diffusion process becomes highly persistent. He used in his simulation 22 years of daily data, assuming 250 business days per year, generated by highly persistent Ornstein–Uhlenbeck processes.

5. Asymptotic distribution of the test if the diffusion is nearly integrated

We now consider the situation where the diffusion, while still formally stationary, is nearly integrated. The idea is similar to what happens in the unit root literature in time series: when studying an AR(1) model with autoregressive parameter ρ, in the nearly-integrated situation where ρ is close to one, it is often the case that the small sample distribution of the parameter estimate of ρ is better approximated by assuming that the process has a unit root (the discrete equivalent to our Brownian motion) than by assuming that the process has ρ close to but strictly smaller than one (in which case the process is stationary).

Here, we derive the asymptotic distribution of the test statistic $\hat{M}$ when the data generating process is nearly integrated, and investigate the ability of that limiting distribution to approximate the small sample behavior of the test when in fact the data generating process is stationary but exhibits very low speeds of mean reversion—as is the case for US interest rates, for instance. More specifically, we consider two interesting classes of diffusions, which are Ornstein–Uhlenbeck processes with T-dependent mean reversion and diffusion parameters. The first class has the mean reversion parameter given as a reciprocal of T, and it approaches Brownian motion as $T \to \infty$. The second class has both the mean reversion and volatility parameters set to be the reciprocals of T, so that its marginal density is stable though it becomes strongly persistent as $T \to \infty$. In what follows, we assume that they have zero means for simplicity, since the presence of non-zero means only introduces a trivial shift effect.

5.1. Near Brownian motion

We let X be generated as

$$dX_t = -\frac{a}{T}X_t dt + \gamma dW_t, \tag{47}$$

i.e., an Ornstein–Uhlenbeck process with mean reversion parameter $\beta = a/T$ for some $a \geq 0$ and diffusion parameter $\gamma > 0$. Clearly, X reduces to a Brownian motion if $a = 0$. Therefore, our analysis here includes the case for which X is a Brownian motion. We easily see that

$$\begin{aligned} X_t &= e^{-(a/T)t}X_0 + \gamma \int_0^t \exp\left[-(a/T)(t-s)\right] dW_s \\ &=_d e^{-a(t/T)}X_0 + \gamma\sqrt{T}\int_0^{t/T} \exp\left[-a(t/T-s)\right] dW_s, \end{aligned}$$

and therefore,

$$X_t =_d \sqrt{T}V^a_{t/T}, \tag{48}$$

where V^a is the Ornstein–Uhlenbeck process with the mean reversion parameter $a > 0$ and diffusion parameter γ. The result in (48) follows in particular from the scale invariant property of the Brownian motion W, i.e., $W_{Ts} =_d \sqrt{T}W_s$. Clearly, V^a reduces to the Brownian motion γW if we set $a = 0$.

We call the diffusion X introduced in (47) the *near* Brownian motion, since it approaches to a Brownian motion as T increases. For the near Brownian motion, we may obtain the limit distribution of the test statistic $\hat{M}$ more explicitly. This is given in the following proposition. We denote by L_a the local time of the Ornstein–Uhlenbeck process V^a.

Proposition 3. *For the near Brownian motion in (47), we have*

$$\frac{\Delta}{h\sqrt{T}}\hat{M} \to_d L_a(1,0) \min_{\theta\in\Theta}\int_{-\infty}^{\infty}\pi^2(x,\theta)w(x)\,dx$$

as $h \to 0$, $\Delta/h^2 \to 0$ and $T \to \infty$.

Of course, the constant term in the limiting distribution of $\hat{M}$ can be computed for a given null parametric family $\pi(\cdot,\theta)$. For example, if we consider the family of normal distributions with parameter $\theta = (\mu,\sigma^2)'$ corresponding to the Ornstein–Uhlenbeck process, then we have that (for $w(x) \equiv 1$)

$$\int_{-\infty}^{\infty}\pi^2(x,\theta)\,dx = \frac{1}{2\sqrt{\pi}\,\sigma}.$$

Therefore, if we set the parameter space for σ^2 to be $[\underline{\sigma}^2, \bar{\sigma}^2]$, then it follows that

$$\min_{\theta\in\Theta}\int_{-\infty}^{\infty}\pi^2(x,\theta)\,dx = \frac{1}{2\sqrt{\pi}\,\bar{\sigma}}.$$

In the Gaussian case, we thus have

$$\hat{M} \approx_d \frac{h\sqrt{T}L_a(1,0)}{2\sqrt{\pi}\,\bar{\sigma}\,\Delta}, \tag{49}$$

and $\hat{M} \to_p \infty$ as $h \to 0$, $\Delta/h^2 \to 0$ and $T \to \infty$.

For the near Brownian motion, the variance of the marginal distribution increases and explodes without a bound as $T \to \infty$. Therefore, the density that most closely approximates the (nonexisting) limiting marginal density of near Brownian motion is naturally given by the most diffuse distribution in the family, i.e., the distribution with the largest variance $\bar{\sigma}^2$ in the case of the normal family. The mean becomes unimportant in this case. In light of this, it is somewhat intuitive that the distribution of $\hat{M}$ given in (49) involves the maximal variance and no mean parameter. For the consistency of the test, the results in Propositions 1 and 2 are applicable with $\kappa = 1/2$ for the near Brownian motion.

5.2. Persistent Ornstein–Uhlenbeck process

Now we consider the diffusion X given by

$$dX_t = -\frac{a}{T}X_t dt + \frac{b}{\sqrt{T}}dW_t, \tag{50}$$

i.e., an Ornstein–Uhlenbeck process with mean reversion parameter $\beta = a/T$ and diffusion parameter $\gamma = b/\sqrt{T}$ for some $a, b > 0$. Both diffusions introduced in (47) and (50) are nearly integrated, in the sense that their associated autoregressive coefficients converge to unity as $T \to \infty$. However, we may easily see that the diffusion X generated as in (50) has the time invariant stationary distribution that is given by normal with mean zero and variance $b^2/2a$ for all T. This is in sharp contrast with the diffusion in (47), whose marginal distribution has variance $\gamma T/2a$ increasing with T. We call the diffusion X generated by (50) the *persistent* Ornstein–Uhlenbeck process, since it is highly persistent and yet has the time invariant marginal distribution of a "regular" Ornstein–Uhlenbeck process.

We may use the persistent Ornstein–Uhlenbeck process to describe a diffusion whose marginal distribution is relatively stable though the process is highly persistent. Such a process was investigated extensively by simulation in Pritsker (1998) for his study of US interest rate models. Below we show that our theory for the persistent Ornstein–Uhlenbeck process is indeed quite useful to understand some of his main simulation findings.

To further analyze the persistent Ornstein–Uhlenbeck process, we note that

$$\begin{aligned} X_t &= e^{-(a/T)t}X_0 + \frac{b}{\sqrt{T}}\int_0^t \exp\left[-(a/T)(t-s)\right] dW_s \\ &=_d e^{-a(t/T)}X_0 + b\int_0^{t/T} \exp\left[-a(t/T-s)\right] dW_s, \end{aligned}$$

from which it follows that

$$X_t =_d V^{a,b}_{t/T}, \tag{51}$$

where $V^{a,b}$ is the Ornstein–Uhlenbeck process with mean reversion parameter $a > 0$ and diffusion parameter $b > 0$. The result in (51) corresponds to and can be derived similarly as (48). In what follows, we signify by $L_{a,b}$ the local time of the Ornstein–Uhlenbeck process $V^{a,b}$.

Lemma 5. *For the persistent Ornstein–Uhlenbeck process in (50), we have*

$$\hat{\pi}(x) \to_d L_{a,b}(1,x)$$

as $h \to 0$, $\Delta/h^2 \to 0$ and $T \to \infty$.

For the persistent Ornstein–Uhlenbeck process, the kernel density estimator $\hat{\pi}$ is not consistent for its time invariant marginal density. This is due to the adverse effect of strong persistence on the kernel estimation. The observations from the persistence Ornstein–Uhlenbeck process do not provide sufficient information about the existing time invariant marginal distribution. As we mentioned above, the persistent Ornstein–Uhlenbeck process has the time invariant marginal distribution given by normal with mean zero and variance $b^2/2a$. As a result, $\hat{\pi}$ remains to be random even in the limit as $T \to \infty$. Needless to say, we may well expect the variance of $\hat{\pi}$ to be excessively large, i.e., much larger than what the asymptotic theory based on stationarity predicts.

However, $\hat{\pi}$ has no asymptotic bias. The limiting distribution of $\hat{\pi}$ is rightfully centered at the true marginal density of the persistent Ornstein–Uhlenbeck process. Note that $L_{a,b}(1,\cdot)$ is a random function, whose expectation is the density of normal distribution with mean zero and variance $b^2/2a$, i.e.,

$$E\left[L_{a,b}(1,x)\right] = \frac{\sqrt{2a}}{b\sqrt{2\pi}}\exp(-ax^2/b^2)$$

for all $-\infty < x < \infty$. Any bias of $\hat{\pi}$ in finite T will therefore vanish as we increase T up to infinity.

Proposition 4. *For the persistent Ornstein–Uhlenbeck process in (50), we have*

$$\frac{\Delta}{hT}\hat{M} \to_d \min_{\theta\in\Theta}\int_{-\infty}^{\infty}(\pi(x,\theta) - L_{a,b}(1,x))^2 L_{a,b}(1,x)w(x)\,dx$$

and

$$\hat{c}(\alpha) \to_d c_1\int_{-\infty}^{\infty}L_{a,b}^2(1,x)w(x)\,dx$$

with the constant c_1 defined in (45), as $h \to 0$, $\Delta/h^2 \to 0$ and $T \to \infty$.

If the bandwidth h is selected according to (46), we have $\hat{M} \to_p \infty$ and $\hat{c}(\alpha) = O_p(1)$ for the persistent Ornstein–Uhlenbeck process. Therefore, the test based on $\hat{M}$ rejects the null of the correct specification with probability approaching to unity as $n = T/\Delta \to \infty$. This implies that the asymptotic size of the test is unity. Recall that the persistent Ornstein–Uhlenbeck process has a proper time invariant marginal distribution, and that the null hypothesis should not be rejected as long as we correctly specify the parametric family π of densities as normal. Here again it is the strong persistency of the persistent Ornstein–Uhlenbeck process, which makes the test reject its true marginal density. Now it becomes rather clear that the test would have excessive empirical sizes for Ornstein–Uhlenbeck processes with small mean reversion and diffusion parameters, which explains the earlier findings on excessive rejection rates observed earlier by Pritsker (1998).

6. Conclusions

The paper explains the behavior of the test statistic of Aït-Sahalia (1996) for diffusions based on the stationary distribution of the process, when the diffusion is either nonstationary or nearly integrated. Because of the key role played by the kernel density estimator, the asymptotic distribution derived under the assumption that the process is mixing is identical to that obtained if the process were i.i.d. As a result, as in the classical near-unit root situation in time series analysis, that distribution can be inaccurate in small samples if the process, while mixing, is close to a unit root. Using a setup that closely mimics the simulation design used by Pritsker (1998), who simulated Ornstein–Uhlenbeck processes with small values of mean reversion and diffusion parameters, we show that our asymptotic theory can explain why the test over-rejects in his simulations.

In the process of analyzing the behavior of that test statistic, we were led to study the asymptotic behavior of the local time, both pointwise and global, and derive results that may hopefully be useful beyond the narrow scope of their application here to analyze the behavior of the marginal density-based test statistic.

Appendix. Proofs

See Appendices A–I.

Appendix A. Proof of Lemma 1

$$\begin{aligned}
&E[L(T,y)\mid X_0=x]\\
&= E\left[\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^T 1\{|X_s-y|<\varepsilon\}ds|X_0=x\right]\\
&= \lim_{\varepsilon\to 0}\int_0^T\frac{1}{2\varepsilon}E\left[1\{|X_s-y|<\varepsilon\}\mid X_0=x\right]ds\\
&= \lim_{\varepsilon\to 0}\int_0^T\frac{1}{2\varepsilon}\int_{y-\varepsilon}^{y+\varepsilon}p(t,z|x)dz\,ds\\
&= \int_0^T p(t,y|x)dt
\end{aligned}$$

and

$$\begin{aligned}
&E\left[L(T,y)^2\mid X_0=x\right]\\
&= E\left[\left(\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^T 1\{|X_s-y|<\varepsilon\}ds\right)^2\middle| X_0=x\right]\\
&= E\left[\left(\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^T 1\{|X_s-y|<\varepsilon\}ds\right)\right.\\
&\quad\times\left.\left(\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^T 1\{|X_t-y|<\varepsilon\}dt\right)\middle| X_0=x\right]\\
&= \lim_{\varepsilon\to 0}\frac{1}{4\varepsilon^2}\int_0^T\int_0^T E\left[1\{|X_s-y|<\varepsilon\}1\right.\\
&\quad\times\left.\{|X_t-y|<\varepsilon\}\mid X_0=x\right]dsdt.
\end{aligned}$$

By symmetry, focus on $s < t$:

$$\begin{aligned}
&\int_0^T\int_0^T E\left[1\{|X_s-y|<\varepsilon\}1\{|X_t-y|<\varepsilon\}\mid X_0=x\right]dsdt\\
&= 2\int_0^T\int_0^t E\left[1\{|X_s-y|<\varepsilon\}1\{|X_t-y|<\varepsilon\}\mid X_0=x\right]dsdt.
\end{aligned}$$

Then by the law of iterated expectations and the Markov property, we have

$$\begin{aligned}
&E\left[1\{|X_s-y|<\varepsilon\}1\{|X_t-y|<\varepsilon\}\mid X_0=x\right]\\
&= E\left[1\{|X_s-y|<\varepsilon\}E\left[1\{|X_t-y|<\varepsilon\}\mid X_s\right]\mid X_0=x\right]\\
&= \int_{y-\varepsilon}^{y+\varepsilon}p(s,z|x)\int_{y-\varepsilon}^{y+\varepsilon}p(t-s,w|z)dwdz
\end{aligned}$$

so that

$$\begin{aligned}
&E\left[L(T,y)^2\mid X_0=x\right]\\
&= 2\lim_{\varepsilon\to 0}\int_0^T\int_0^t\left(\frac{1}{4\varepsilon^2}\int_{y-\varepsilon}^{y+\varepsilon}p(s,z|x)\right.\\
&\quad\times\left.\int_{y-\varepsilon}^{y+\varepsilon}p(t-s,w|z)dwdz\right)dsdt\\
&= 2\int_0^T\int_0^t p(s,y|x)p(t-s,y|y)dsdt.
\end{aligned}$$

Appendix B. Proof of Lemma 2

Let $C_T = T^{1/(r+2)} = T^{1-\kappa}$, and define

$$X_t^T = C_T^{-1}X_{Tt} \quad\text{and}\quad W_t^T = T^{-1/2}W_{Tt}.$$

Then we may easily deduce that

$$dX_t^T = c\,|X_t^T|^{-r/2}dW_t^T.$$

Therefore, it follows that

$$\left(X_t^T\right) =_d \left(C_T^{-1}X_{Tt}\right) \tag{B.1}$$

for all $T > 0$.

Consequently, we have

$$\begin{aligned}\frac{1}{2\varepsilon}\int_0^T 1\{|X_t - x| < \varepsilon\}\,dx &= \frac{T}{2\varepsilon}\int_0^1 1\{|X_{Tt} - x| < \varepsilon\}\,dx \\ &=_d \frac{T}{2\varepsilon}\int_0^1 1\{C_T|X_t - C_T^{-1}x| < \varepsilon\}\,dx \\ &= \frac{T}{C_T}\frac{C_T}{2\varepsilon}\int_0^1 1\{|X_t - C_T^{-1}x| < C_T^{-1}\varepsilon\}\,dx \qquad \text{(B.2)}\end{aligned}$$

due in particular to (B.1). The stated result now follows immediately from (B.2) by taking the limit $\varepsilon \to 0$ and noting $T/C_T = T^{\kappa}$.

Appendix C. Proof of Lemma 3

Let

$$\Delta\sum_{i=1}^n f(X_{i\Delta}) = \Delta\sum_{i=1}^n f(X_{(i-1)\Delta}) + R(T,\Delta)$$

with $R(T,\Delta) = \Delta[f(X_T) - f(X_0)]$, and subsequently

$$\Delta\sum_{i=1}^n f(X_{(i-1)\Delta}) = \int_0^T f(X_t)\,dt - S(T,\Delta)$$

with

$$S(T,\Delta) = \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} [f(X_t) - f(X_{(i-1)\Delta})]\,dt. \qquad \text{(C.3)}$$

Now it suffices to show that $R(T,\Delta)$ and $S(T,\Delta)$ are of order smaller than $\int_0^T f(X_t)\,dt$. It is trivial to see that this is true for $R(T,\Delta)$.

To analyze $S(T,\Delta)$ in (C.3), we first let $\mathcal{D}$ be the differential operator, and define the infinitesimal generator of the diffusion X as the operator

$$\mathcal{A} = \mu\mathcal{D} + \frac{1}{2}\sigma^2\mathcal{D}^2, \qquad \mathcal{B} = \sigma\mathcal{D}.$$

If we define $f_A = \mathcal{A}f$ and $f_B = \mathcal{B}f$ for a twice continuously differentiable function f, then $\int_0^t f_A(X_s)\,ds$ and $\int_0^t f_B(X_s)\,dW_s$ represent, respectively, the bounded variation and martingale parts of $f(X_t) - f(X_0)$, or in differential notation

$$df(X_t) = f_A(X_t)\,dt + f_B(X_t)\,dW_t, \qquad \text{(C.4)}$$

which follows from Itô's formula.

We now write using (C.4)

$$f(X_t) - f(X_{(i-1)\Delta}) = \int_{(i-1)\Delta}^t f_A(X_s)\,ds + \int_{(i-1)\Delta}^t f_B(X_s)\,dW_s,$$

and deduce that

$$\begin{aligned}\int_{(i-1)\Delta}^{i\Delta} [f(X_t) - f(X_{(i-1)\Delta})]\,dt &= \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^t f_A(X_s)\,ds\,dt + \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^t f_B(X_s)\,dW_s\,dt \\ &= \int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)f_A(X_t)\,dt + \int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)f_B(X_t)\,dW_t\end{aligned}$$

by changing the order of integration. It follows that

$$S(T,\Delta) = S_1(T,\Delta) + S_2(T,\Delta),$$

where

$$S_1(T,\Delta) = \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)f_A(X_t)\,dt \quad \text{and}$$

$$S_2(T,\Delta) = \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)f_B(X_t)\,dW_t.$$

Clearly,

$$|S_1(T,\Delta)| \le \Delta\int_0^T |f_A(X_t)|\,dt. \qquad \text{(C.5)}$$

Moreover, if we denote by $[S_2](T,\Delta)$ the quadratic variation of $S_2(T,\Delta)$, we have

$$[S_2](T,\Delta) \le \Delta^2\int_0^T f_B^2(X_t)\,dt. \qquad \text{(C.6)}$$

The order of $S_2(T,\Delta)$ in probability is given by the square root of $[S_2](T,\Delta)$.

First, we let $f = a_p$. Under the given conditions, we have $|f_A(x)| \le c|x|^{r+p-1}$ for some constant c. We therefore have from (C.5)

$$S_1(T,\Delta) \sim \Delta T^{(r+p)(1-\kappa)+\kappa} \sim \Delta T^{(r-1)(1-\kappa)}\int_0^T f(X_t)\,dt.$$

Moreover, it follows from (C.6) that

$$S_2(T,\Delta) \sim \Delta T^{[(r+2p)(1-\kappa)+\kappa]/2},$$

which is of order smaller than $S_1(T,\Delta)$ in probability. Next, we let $f = b$. Under the given conditions, both f_A and f_B are bounded and integrable. We may therefore easily deduce that

$$S_1(T,\Delta), S_2(T,\Delta) \sim \Delta\int_0^T f(X_t)\,dt$$

due to (C.5) and (C.6). The proof is therefore complete.

Appendix D. Proof of Lemma 4

The proof heavily relies on the results in Aït-Sahalia and Park (2009), which will be referred to as AP hereafter. The proof of the first part is given by Theorem 1 in AP. To derive the second part, we define

$$\begin{aligned}\tilde{L}(T,x) &= \frac{1}{h}\int_0^T K\left(\frac{X_t - x}{h}\right)dt \\ &= \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{u - x}{h}\right)L(T,u)\,du \\ &= \int_{-\infty}^{\infty} K(u)L(T,x+hu)\,du\end{aligned}$$

and write

$$\hat{L}(T,x) - L(T,x) = [\hat{L}(T,x) - \tilde{L}(T,x)] + [\tilde{L}(T,x) - L(T,x)], \qquad \text{(D.7)}$$

each of which will be looked at subsequently below.

Due to Lemma A2 in AP, we may bound

$$\tilde{L}(T,x) - L(T,x) = \int_{-\infty}^{\infty} K(u)[L(T,x+hu) - L(T,x)]\,du$$

by a constant multiple of $m(x)h^2L(T,x)$ uniformly in $x \in D$. Consequently, we have

$$\int_{-\infty}^{\infty} a_p(x)[\tilde{L}^q(T,x) - L^q(T,x)]\,dx$$
$$\approx c_q h^2 \int_{-\infty}^{\infty} a_p(x)m(x)L^q(T,x)\,dx$$
$$\sim h^2 T^{r_1(1-\kappa)} \int_{-\infty}^{\infty} a_p(x)L^q(T,x)\,dx = o(1)$$
$$\times \int_{-\infty}^{\infty} a_p(x)L^q(T,x)\,dx \quad \text{a.s.}$$

and

$$\int_{-\infty}^{\infty} b(x)[\tilde{L}^q(T,x) - L^q(T,x)]\,dx$$
$$\approx c_q h^2 \int_{-\infty}^{\infty} b(x)m(x)L^q(T,x)\,dx$$
$$\sim h^2 \int_{-\infty}^{\infty} b(x)L^q(T,x)\,dx = o(1)\int_{-\infty}^{\infty} b(x)L^q(T,x)\,dx \quad \text{a.s.,}$$

where c_q is a constant depending only upon q. We may therefore easily deduce that

$$\int_{-\infty}^{\infty} f(x)\tilde{L}^q(T,x)\,dx \approx \int_{-\infty}^{\infty} f(x)L^q(T,x)\,dx \tag{D.8}$$

for both $f = a_p$ and $f = b$.

We may similarly analyze

$$\hat{L}(T,x) - \tilde{L}(T,x) = \frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right) - \frac{1}{h}\int_0^T K\left(\frac{X_t - x}{h}\right)dt,$$

which is bounded uniformly in $x \in D$ by a constant multiple of

$$\left(\Delta h^{-1}|\mu(x)| + \Delta h^{-2}|\sigma^2|\right)L(T,x),$$

as shown in Lemma A3 of AP. Indeed, we may easily show precisely as above that

$$\int_{-\infty}^{\infty} f(x)[\hat{L}^q(T,x) - \tilde{L}^q(T,x)]\,dx \sim o(1)\int_{-\infty}^{\infty} f(x)L^q(T,x)\,dx \quad \text{a.s.} \tag{D.9}$$

for both $f = a_p$ and $f = b$. The stated result now follows immediately from (D.7) to (D.9).

Appendix E. Proof of Proposition 1

Due to Assumption 3 and Lemma 3, we have

$$\int_{-\infty}^{\infty} \hat{\pi}^r(x)\,dx \sim T^{-r}\int_{-\infty}^{\infty} L^r(T,x)\,dx \sim T^{-(1-\kappa)(r-1)} \tag{E.10}$$

and

$$\int_{-\infty}^{\infty} f(x)\hat{\pi}^r(x)\,dx \sim T^{-r}\int_{-\infty}^{\infty} f(x)L^r(T,x)\,dx \sim T^{-(1-\kappa)r} \tag{E.11}$$

for any bounded integrable f. Moreover, we have

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_{i\Delta}^2 \approx T^{-1}\int_0^T X_t^2 dt = T^{-1}$$
$$\times \int_{-\infty}^{\infty} x^2 L(T,x)\,dx \sim T^{2(1-\kappa)}, \tag{E.12}$$

due to Assumption 3.

Let $w(x) \equiv 1$. We may now easily deduce from (E.10) and (E.11) that

$$\int_{-\infty}^{\infty} (\pi(x,\theta) - \hat{\pi}(x))^2\hat{\pi}(x)\,dx = \int_{-\infty}^{\infty} \pi^2(x,\theta)\hat{\pi}(x)\,dx$$
$$- 2\int_{-\infty}^{\infty} \pi(x,\theta)\hat{\pi}^2(x)\,dx + \int_{-\infty}^{\infty} \hat{\pi}^3(x)\,dx$$

is of order in probability

$$T^{-(1-\kappa)} + T^{-2(1-\kappa)} + T^{-2(1-\kappa)} \sim T^{-(1-\kappa)}.$$

In particular, the order is given uniformly in $\theta \in \Theta$, due to Assumption 5 and continuity of

$$\int_{-\infty}^{\infty} \pi^2(x,\theta)L(T,x)\,dx \quad \text{and} \quad \int_{-\infty}^{\infty} \pi(x,\theta)L^2(T,x)\,dx$$

with respect to $\theta \in \Theta$. This can be shown using the standard argument to establish the uniform convergence for a sequence of continuous random functions defined on a compact set, based on constructing localized supremum and infimum functions $\sup_{\theta\in N(\theta)}\pi(\cdot,\theta)$ and $\inf_{\theta\in N(\theta)}\pi(\cdot,\theta)$ for each point $\theta \in \Theta$ with a neighborhood $N(\theta)$, and extracting using compactness a finite number of such pairs of functions which collapse down to $\pi(\cdot,\theta)$ in the limit for all $\theta \in \Theta$.

Consequently, we have

$$\hat{M} \sim nhT^{-(1-\kappa)} \sim h\Delta^{-1}T^{\kappa}. \tag{E.13}$$

On the other hand,

$$\hat{E}_M \sim T^{-(1-\kappa)} \quad \text{and} \quad \hat{V}_M \sim T^{-3(1-\kappa)}.$$

Therefore,

$$\hat{c}(\alpha) \sim T^{-(1-\kappa)} + h^{1/2}T^{3(1-\kappa)/2}, \tag{E.14}$$

as one may easily see.

The test rejects the correct specification if

$$\hat{c}(\alpha) = o_p(\hat{M}),$$

which holds if

$$T^{-(1-\kappa)} + h^{1/2}T^{3(1-\kappa)/2} = o(h\Delta^{-1}T^{\kappa}),$$

due to (E.13) and (E.14). This is in turn satisfied if the stated condition holds. Furthermore, if h is selected as in (46) and given by

$$h \sim T^{7/9-\kappa}\Delta^{2/9}(\ln(T/\Delta))^{-1} \tag{E.15}$$

due in particular to (E.12), then we have

$$hT \sim T^{16/9-\kappa}\Delta^{2/9}(\ln(T/\Delta))^{-1} \quad \text{and}$$
$$h^{1/2}T^{(5\kappa-3)/2} \sim T^{2\kappa-10/9}\Delta^{1/9}(\ln(T/\Delta))^{-1/2}.$$

Therefore, the condition given in the first part is equivalent to

$$\Delta^{7/9} = o(T^{16/9-\kappa}(\ln(T/\Delta))^{-1}) \quad \text{and}$$
$$\Delta^{8/9} = o(T^{2\kappa-10/9}(\ln(T/\Delta))^{-1/2}),$$

i.e.,

$$\Delta = o(T^{(16-9\kappa)/7}(\ln(T/\Delta))^{-9/7}) \quad \text{and}$$
$$\Delta = o(T^{(9\kappa-5)/4}(\ln(T/\Delta))^{-9/16}),$$

which would obviously hold if we have the condition stated in the proposition. Note that $(16-9\kappa)/7 \geq (9\kappa-5)/4$ for all $\kappa \leq 1$.

Appendix F. Proof of Proposition 2

If we set $w(x) = 1\{|x| \le c\}$ for some constant c, then we have from (E.11)

$$\hat{M} \sim nhT^{-(1-\kappa)} \sim h\Delta^{-1}T^{\kappa}, \tag{F.16}$$

as in the proof of Proposition 1. However, we have

$$\hat{E}_M \sim T^{-2(1-\kappa)} \quad \text{and} \quad \hat{V}_M \sim T^{-4(1-\kappa)}.$$

Therefore,

$$\hat{c}(\alpha) \sim T^{-2(1-\kappa)} + h^{1/2}T^{2(1-\kappa)}, \tag{F.17}$$

which differs from the previous case.

Now the test rejects the correct specification if

$$T^{-2(1-\kappa)} + h^{1/2}T^{2(1-\kappa)} = o(h\Delta^{-1}T^{\kappa}),$$

as follow from (F.16) and (F.17), which in turn holds when the stated condition holds. If h is selected as in (46), then we have from (E.15)

$$hT^{2-\kappa} \sim T^{25/9-2\kappa}\Delta^{2/9}(\ln(T/\Delta))^{-1} \quad \text{and}$$
$$h^{1/2}T^{3\kappa-2} \sim T^{5\kappa/2-29/18}\Delta^{1/9}(\ln(T/\Delta))^{-1/2}.$$

Therefore, the condition given in the first part is equivalent to

$$\Delta^{7/9} = o(T^{25/9-2\kappa}(\ln(T/\Delta))^{-1}) \quad \text{and}$$
$$\Delta^{8/9} = o(T^{5\kappa/2-29/18}(\ln(T/\Delta))^{-1/2}),$$

i.e.,

$$\Delta = o(T^{(25-18\kappa)/7}(\ln(T/\Delta))^{-9/7}) \quad \text{and}$$
$$\Delta = o(T^{(45\kappa-29)/16}(\ln(T/\Delta))^{-9/16}),$$

which would obviously hold if we have the condition stated in the proposition. Note that $(25-18\kappa)/7 \ge (45\kappa-29)/16$ for all $\kappa \le 1$.

Appendix G. Proof of Proposition 3

As earlier, we let L be the local time of X. We have

$$\begin{aligned}
&\frac{1}{2\varepsilon}\int_0^T 1\{|X_t - x| < \varepsilon\}dt \\
&=_d \frac{1}{2\varepsilon}\int_0^T 1\left\{\left|\sqrt{T}V^a_{t/T} - x\right| < \varepsilon\right\} dt \\
&= \frac{T}{2\varepsilon}\int_0^1 1\left\{\left|\sqrt{T}V^a_t - x\right| < \varepsilon\right\} dt \\
&= \sqrt{T}\frac{\sqrt{T}}{2\varepsilon}\int_0^1 1\left\{\left|V^a_t - \frac{x}{\sqrt{T}}\right| < \frac{\varepsilon}{\sqrt{T}}\right\} dt.
\end{aligned}$$

Therefore, we obtain

$$L(T,x) =_d \sqrt{T}L_a\left(1, \frac{x}{\sqrt{T}}\right) \tag{G.18}$$

by taking the limit $\varepsilon \to 0$.

Since $L(T,x) = O_p(T^{1/2})$, it follows from Theorem 1 in AP that

$$\hat{L}(T,x) = L(T,x) + O_p(\Delta h^{-2}T^{1/2}) + O_p(h^{1/2}T^{1/4}) + O_p(h^2T^{1/2}).$$

Therefore we have, due to (G.18),

$$\begin{aligned}
\sqrt{T}\hat{\pi}(x) &= \frac{\hat{L}(T,x)}{\sqrt{T}} \\
&= \frac{L(T,x)}{\sqrt{T}} + O_p(\Delta h^{-2}) \\
&\quad + O_p(h^{1/2}T^{-1/4}) + O_p(h^2) \\
&=_d L_a\left(1, \frac{x}{\sqrt{T}}\right) + o_p(1),
\end{aligned}$$

if $h \to 0$, $\Delta/h^2 \to 0$ and $T \to \infty$.

Now we may easily deduce that

$$\sqrt{T}\int_{-\infty}^{\infty}\pi^2(x,\theta)\hat{\pi}(x)w(x)\,dx \to_d L_a(1,0)\int_{-\infty}^{\infty}\pi^2(x,\theta)w(x)\,dx \tag{G.19}$$

uniformly in $\theta \in \Theta$. Likewise, we have

$$T\int_{-\infty}^{\infty}\pi(x,\theta)\hat{\pi}^2(x)w(x)\,dx \to_d L_a^2(1,0)\int_{-\infty}^{\infty}\pi(x,\theta)w(x)\,dx \tag{G.20}$$

uniformly in $\theta \in \Theta$. Finally, it follows that

$$\begin{aligned}
T\int_{-\infty}^{\infty}\hat{\pi}^3(x)\,dx &=_d \frac{1}{\sqrt{T}}\int_{-\infty}^{\infty}L_a^3\left(1, \frac{x}{\sqrt{T}}\right)dx \\
&\quad + o_p(1) \to_p \int_{-\infty}^{\infty}L_a^3(1,x)\,dx
\end{aligned} \tag{G.21}$$

and

$$\begin{aligned}
&T^{3/2}\int_{-\infty}^{\infty}\hat{\pi}^3(x)1\{|x| \le c\}\,dx \to_d L_a^3(1,0) \\
&\quad \times \int_{-\infty}^{\infty}1\{|x| \le c\}\,dx = 2cL_a^3(1,0),
\end{aligned} \tag{G.22}$$

respectively for the cases of $w(x) \equiv 1$ and $w(x) = 1\{|x| \le c\}$ for some $c > 0$. The asymptotics for $\hat{M}$ can now be derived readily from (G.19)–(G.22).

Appendix H. Proof of Lemma 5

The proof is analogous to that of Proposition 3. We have

$$\begin{aligned}
\frac{1}{2\varepsilon}\int_0^T 1\{|X_t - x| < \varepsilon\}dt &=_d \frac{1}{2\varepsilon}\int_0^T 1\{|V^{a,b}_{t/T} - x| < \varepsilon\}dt \\
&= \frac{T}{2\varepsilon}\int_0^1 1\{|V^{a,b}_t - x| < \varepsilon\}dt,
\end{aligned}$$

and therefore, we obtain

$$L(T,x) =_d TL_{a,b}(1,x) \tag{H.23}$$

by taking the limit $\varepsilon \to 0$. Now we have $L(T,x) = O_p(T)$, and may deduce from Theorem 1 in AP that

$$\hat{L}(T,x) = L(T,x) + O_p(\Delta h^{-2}T) + O_p(h^{1/2}T^{1/2}) + O_p(h^2T). \tag{H.24}$$

Consequently, it follows from (H.23) and (H.24) that

$$\begin{aligned}
\hat{\pi}(x) &= \frac{\hat{L}(T,x)}{T} \\
&= \frac{L(T,x)}{T} + O_p(\Delta h^{-2}) + O_p(h^{1/2}T^{-1/2}) + O_p(h^2) \\
&=_d L_{a,b}(1,x) + o_p(1),
\end{aligned}$$

if $h \to 0$, $\Delta/h^2 \to 0$ and $T \to \infty$. This was to be shown.

Appendix I. Proof of Proposition 4

We may readily deduce from Lemma 5 that

$$\begin{aligned}
&\int_{-\infty}^{\infty}(\pi(x,\theta) - \hat{\pi}(x))^2\hat{\pi}(x)w(x)\,dx \\
&\to_d \int_{-\infty}^{\infty}(\pi(x,\theta) - L_{a,b}(1,x))^2L_{a,b}(1,x)w(x)\,dx
\end{aligned} \tag{I.25}$$

uniformly in $\theta \in \Theta$, and that

$$\hat{E}_M \to_d c_1 \int_{-\infty}^{\infty} L_{a,b}^2(1,x)w(x)\,dx \tag{I.26}$$

$$\hat{V}_M \to_d c_2 \int_{-\infty}^{\infty} L_{a,b}^4(1,x)w(x)\,dx. \tag{I.27}$$

Moreover, we have

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} X_{i\Delta}^2 \approx T^{-1}\int_0^T X_t^2 dt \\ &= T^{-1}\int_{-\infty}^{\infty} x^2 L(T,x)\,dx =_d \int_{-\infty}^{\infty} x^2 L_{a,b}(1,x)\,dx,\end{aligned} \tag{I.28}$$

due in particular to (H.23). The stated results now follow immediately from (I.25) to (I.28).

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385–426.

Aït-Sahalia, Y., Park, J., 2009. Nonparametric estimation of non-stationary diffusion models. Tech. Rep. Princeton University and Indiana University.

Akonom, J., 1993. Comportement asymptotique du temps d'occupation du processus des sommes partielles. Annales de l'Institut Henri Poincaré 29, 57–81.

Bandi, F.M., Moloche, G., 2005. On the functional estimation of multivariate diffusion processes. Tech. Rep. University of Chicago.

Bandi, F.M., Phillips, P.C.B., 2003. Fully nonparametric estimation of scalar diffusion models. Econometrica 71, 241–283.

Bandi, F.M., Phillips, P.C.B., 2009. Nonstationary continuous-time processes. In: Aït-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics. North Holland, Amsterdam, The Netherlands.

Bosq, D., 1998. Nonparametric Statistics for Stochastic Processes, second ed. In: Lecture Notes in Statistics, vol. 110. Springer, New York.

Chapman, D.A., Pearson, N., 2000. Is the short rate drift actually nonlinear? Journal of Finance 55, 355–388.

Chen, S.X., Gao, J., 2004a. An adaptive empirical likelihood test for time series models. Tech. Rep. Iowa State University.

Chen, S.X., Gao, J., 2004b. On the use of the kernel method for specification tests of diffusion models. Tech. Rep. Iowa State University.

Corradi, V., Swanson, N.R., 2005. A bootstrap specification test for diffusion processes. Journal of Econometrics 124, 117–148.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–408.

Darling, D.A., Kac, M., 1957. On occupation times for Markov processes. Transactions of the American Mathematical Society 84, 444–458.

Ethier, S.N., Kurtz, T.G., 1986. Markov Processes: Characterization and Convergence. John Wiley and Sons, New York.

Fan, J., Zhang, C., 2003. A reexamination of diffusion estimations with applications to financial model validation. Journal of American Statistical Association 13, 965–992.

Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Review of Financial Studies 18, 37–84.

Höpfner, R., Löcherbach, E., 2003. Limit Theorems for Null Recurrent Markov Processes. In: Memoirs of the American Mathematical Society, vol. 161. American Mathematical Society, Providence, RI.

Karatzas, I., Shreve, S.E., 1991. Brownian Motion and Stochastic Calculus. Springer-Verlag, New York.

Karlin, S., Taylor, H.M., 1981. A Second Course in Stochastic Processes. Academic Press, New York.

Li, F., Tkacz, G., 2006. A consistent bootstrap test for conditional density functions with time-series data. Journal of Econometrics 133, 863–886.

Park, J.Y., 2007, The spatial analysis of time series. Tech. Rep. Indiana University.

Pritsker, M., 1998. Nonparametric density estimation and tests of continuous time interest rate models. Review of Financial Studies 11, 449–487.

Revuz, D., Yor, M., 1994. Continuous Martingales and Brownian Motion, second ed. Springer-Verlag, Berlin, Germany.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.