A Nonparametric Test for Stationarity in Continuous-Time Markov Processes*

Shin Kanaya†

Department of Economics, Nuffield College and Oxford-Man Institute, University of Oxford

Job Market Paper - November 2011

Abstract

In this paper, we propose a new nonparametric testing procedure to examine the stationarity property of an underlying continuous-time Markov process. The stationarity is often assumed in building/estimating dynamic models in economics and finance. However, existing statistical methods to check the stationarity typically rely on a particular parametric assumption called a unit root. The unit-root concept is well defined for a certain class of parametric models in discrete time settings (e.g., linear auto-regression models with finite-variance error disturbances) but not necessarily for general nonlinear models and/or continuous-time models. To check the stationarity property, we exploit a restriction implied by the infinitesimal generator - a functional operator computed via the derivatives of conditional expectations with respect to time. This restriction allows us to develop a new theorem for identifying the *generic* stationarity property fully nonparametrically within a class of univariate time-homogeneous Markov processes. We construct a kernel-based test statistic based on this theorem, and derive its null asymptotic distribution. We also prove that the proposed test is consistent against nonstationary (null recurrent) processes. Our proofs for the asymptotic results proceed by using the so-called regeneration and ratio-limit properties of Markov processes without imposing any type of mixing condition. We conduct Monte-Carlo simulations to study finite-sample size and power properties of the test, and apply the proposed method to foreign exchange rates and short-term interest rates to assess the validity of the stationarity hypothesis.

*The author wishes to thank Federico M. Bandi, Marine Carrasco, Yoosoon Chang, Juan Carlos Escanciano, Jean-Pierre Florens, Nikolay Gospodinov, Bruce E. Hansen, Ilze Kalnina, Toru Kitagawa, Dennis Kristensen, William McCausland, Bent Nielsen, Joon Y. Park, Benoit Perron, Jack R. Porter, Neil Shephard, Xiaoxiao Shi and seminar participants at University of Wisconsin-Madison, Indiana University, University of Montreal, the 5th CIREQ Time Series Conference, and the 2011 Midwest Econometrics Group Conference for helpful comments and suggestions.

†Address: The Oxford-Man Institute, Eagle House, Walton Well Road, Oxford, OX2 6ED, UK. E-mail: shin.kanaya@economics.ox.ac.uk. Phone: +44 (0) 1865-616637. Fax: +44 (0) 1865-616601.

1 Introduction

In this paper, we propose a new nonparametric testing procedure to examine if an underlying continuous-time process is stationary/stable within a class of univariate time-homogeneous Markov processes. The stationarity is often assumed in constructing/estimating dynamic models in economics and finance. However, most testing methods to discriminate stationarity and nonstationarity rely on the concepts of a *unit root* or *integration*, such as the Dickey-Fuller and KPSS type tests (Dickey and Fuller, 1979, and Kwiatkowski, Phillips, Schmidt and Shin 1992, respectively). These concepts are well defined for linear models in the discrete-time framework (in particular, linear autoregressive models with finite-variance disturbances); however, not necessarily for general nonlinear models and/or continuous-time models. As a result, many of the existing tests based on the unit root or integration concept may not be useful to examine the *generic* stationarity/stability property of time-series processes.

As a concrete example, consider the case where DF-type tests (whose null is the nonstationary unit root hypothesis) are applied to so-called stochastic unit-root (STUR) processes (introduced in Granger and Swanson, 1997). A STUR process can be stationary or nonstationary, depending on its parameter setting. If the data-generating process is a stationary STUR process, the DF-type tests do not often lead to a rejection result (Granger and Swanson, 1997). On the other hand, in the case of a nonstationary STUR process, they lead to a rejection result (Nagakura, 2009). That is, if we use the DF-type tests to check the stationarity/nonstationarity, they are likely to give us an opposite conclusion that is wrong. This is crucial when tying some economic theory directly to the stationarity concept (say, the purchasing-power-parity hypothesis or the law of one price in international economics), as the DF-type tests cannot appropriately examine the empirical validity of such economic theory. Note that STUR processes are defined in the discrete-time framework, and may not be necessarily suited to our continuous-time framework. However, the problem here is that the stationarity/nonstationarity property of general nonlinear and/or continuous-time processes has nothing to do with the unit root concept, and such processes are not in the scope of the DF type tests. We emphasize that the unit root represents only one of the possible forms of nonstationarity, although its importance in econometric modeling is inarguable.

We also note that traditional DF and KPSS type tests focus only on the (so-called) *drift-induced stationarity*, in which the form of the drift (conditional-mean) function ensures stationarity. They may not necessarily exploit volatility information to examine the stationarity property. As argued in the financial econometrics literature (e.g., Conley, Hansen, Lutter and Scheinkman, 1997, Nicolau, 2005), the stationarity may be *volatility-induced*. Processes with the volatility-induced stationarity, to which the unit-root or integration concept is not applicable and whose variances are often infinite, are not generally in the scope of unit-root and/or KPSS type tests.[1]

In this paper, we construct a test to examine the *generic* stationarity property. This is possible by

[1] Indeed, the stationarity in STUR processes exemplified above may be interpreted as being volatility-induced.

exploiting a restriction for the stationarity implied by the infinitesimal generator - a functional operator computed via the derivatives of conditional expectations with respect to time. This restriction does not rely on particular forms of conditional expectation and volatility functions and allows us to develop a new theorem for identifying the stationarity property fully nonparametrically. We construct a kernel-based test statistic based on this theorem, and derive its asymptotic null distribution. We also prove that the proposed test is consistent against nonstationary (null recurrent) processes. Our proofs for the asymptotic results proceed by using the so-called regeneration and ratio-limit properties of Markov processes. We do not impose any type of mixing condition (or some other weak-dependence condition) to establish distributional theory. We note that for the purpose of our analysis, it is important to work without any mixing condition. To construct a statistical testing procedure, it is generally reasonable to maintain the same conditions under both the null and alternative hypotheses. If we imposed some mixing condition for both hypotheses, we would have the class of alternative processes essentially empty. While the mixing is in principle a different concept from the stationarity (e.g., there exist some stationary processes which do not satisfy a certain mixing condition), they are quite interrelated. We also conduct Monte-Carlo simulations to study finite-sample size and power properties of the test, and apply the proposed method to foreign exchange rates and short-term interest rates to examine the validity of the stationarity hypothesis.

Recently, Bandi and Corradi (2011) have proposed nonparametric tests to check the null hypothesis of nonstationarity for Markov processes. It is known that a sort of the law of large numbers (LLN) holds for any recurrent Markov process (due to the Markov regeneration), but its convergence rate in the stationary case is different from that in the nonstationary case.[2] Bandi and Corradi exploited this difference to develop nonstationary tests. While their tests seem to be intended mainly for discrete-time processes, they are also applicable to some (limited) class of continuous-time processes (diffusion processes). In contrast, we only work with continuous-time processes, but more general processes (beyond diffusion processes) are within the scope of our test. In this respect, their tests and ours complement each other. On the other hand, one needs to specify the LLN-convergence rate of a process to define the null hypothesis and construct Bandi and Corradi's test statistic. While such a rate is generally unknown, it seems to crucially determine the properties of their tests. In a related study, Berenguer-Rico and Gonzalo (2011) proposed a generalization of the integration concept, called summability. They present a method to estimate the degree of summability, which indeed corresponds to the LLN-convergence rate of a process. Berenguer-Rico and Gonzalo's method might also be used to conduct a formal statistical test for stationarity/nonstationarity (upon developing some distributional theory).

What distinguishes this paper from these two papers is that we do not directly use the restriction of the convergence rate in the (generalized) LLN, but instead use the restriction based on infinitesimal generators. For infinitesimal generators to be well defined, it is crucial to maintain the Markov assump-

[2] Roughly, it holds that $n^{-d}\sum_{i=1}^{n} X_i = O_P(1)$ with $d = 1$ if a Markov sequence $\{X_i\}_{i=1}^{n}$ is stationary, but with some $d \in (0, 1)$ if it is nonstationary (null recurrent).

tion. In view of this, the Markov assumption plays a more important role in our paper than in Bandi and Corradi's (2011) paper. We note that the Markov property itself is also an interesting property to be examined. Several authors (Aït-Sahalia, Fan and Jiang, 2010; Amaro de Matos and Fernandes, 2007; Chen and Hong, 2011) have developed testing procedures to check the Markov property. However, their tests presuppose stationarity and weak time-series dependence of processes; therefore, they cannot be used as pretests for our Markov assumption. At the same time, our test, which relies on the Markov assumption, cannot be used as a pretest for their stationarity assumption. In view of this, existing Markov tests and ours are not directly related, but may be regarded as complementary to each other in terms of checking the validity of Markov and stationary assumptions, which are often maintained in practical time-series modeling.

The rest of the paper is organized as follows. The next section describes our framework, introducing continuous-time Markov processes and the corresponding infinitesimal generators with some examples. We also clarify technical requirements, which should be imposed on the Markov processes. Section 3 presents our identification theorem for the stationarity property. In section 4, we propose our test statistic and investigate its asymptotic behavior. Section 6 provides some concluding remarks. All proofs can be found in the Appendix.

We use the following notation throughout the text: $g'(x)$, $g''(x)$, $g'''(x)$ and $g^{(k)}(x)$ denote the first, second, third, and k-th order derivatives of a function g. The symbols $\stackrel{P}{\to}$ and $\Longrightarrow$ mean convergence in probability and weak convergence, respectively. For definitional equations, we use the notations: $A := B$ and $C =: D$. The former means that A is defined by B, and the latter means that D is defined by C.

2 Framework

Here, we describe our basic setup and introduce infinitesimal generators. Let $\{X_s\} := \{X_s\}_{s\geq 0}$ be a scalar, time-homogeneous and continuous-time Markov process defined on a filtered probability space $\left(\Omega, \mathfrak{F}, \{\mathfrak{F}_s\}_{s\geq 0}, \Pr\right)$, which satisfies the usual conditions. Let I denote the state space of $\{X_s\}$. For simplicity, we consider the case where I is the whole real line $\mathbb{R} := (-\infty, \infty)$. We denote by $\mathfrak{B}(I)$ the Borel algebra on I.[3] The time-homogeneous Markov process is determined by the transition function $P(s, x, \Gamma)$ and the (initial) distribution of X_0. $P(s, x, \Gamma)$ represents $\Pr[X_s \in \Gamma | X_0 = x]$, the probability that the process which has started from point $x \in I$ is in the set $\Gamma \in \mathfrak{B}(I)$ at time $s \in [0, \infty)$. Time-homogeneity means that $\Pr[X_s \in \Gamma | X_0 = x] = \Pr[X_{s+t} \in \Gamma | X_t = x]$ for any $s, t \in [0, \infty)$. Throughout the paper, we call $P : [0, \infty) \times I \times \mathfrak{B}(I) \to [0, 1]$ as a transition function when it satisfies the following conditions:

[3] The topology is generated by the usual Euclidean norm.

Assumption 1 *i) For any $s \in [0,\infty)$ and $x \in I$, $P(s,x,\cdot)$ is a probability measure on $\mathfrak{B}(I)$; and for any $\Gamma \in \mathfrak{B}(I)$, $P(\cdot,\cdot,\Gamma)$ is $\mathfrak{B}(\mathbb{R}) \times \mathfrak{B}(I)$-measurable, where $\mathfrak{B}(\mathbb{R})$ is the Borel algebra on $\mathbb{R}$. ii) For any $s,t \in [0,\infty)$, $P(s+t,x,\Gamma) = \int_I P(s,x,dy)\, P(t,y,\Gamma)$. iii) For any $x \in I$ and an arbitrary neighborhood U of x, $P(s,x,U) \to 1$ as $s \to 0$.*

These conditions in Assumption 1 are quite standard when considering Markov processes (see, e.g., Ch. 2 of Dynkin, 1965). By the condition (i), $P(s,x,I) = 1$ for any $s \in [0,\infty)$ and $x \in I$. This is called the conservativeness condition, meaning that there is no (isolated) coffin state where the process is killed/terminated and the process always remains somewhere in I.[4] The condition (ii) is called as the Chapman-Kolmogorov condition, and (iii) is often referred to as the stochastic continuity condition.

Given the transition function, we now define a functional operator. Let $\mathbf{B}(I)$ denote the Banach space of all $\mathfrak{B}(I)$-measurable bounded functions on I with the sup-norm $||f|| := \sup_{x\in I} |f(x)|$. For each $s \geq 0$, define a functional operator on $\mathbf{B}(I)$ as

$$\mathcal{T}_s \varphi(x) := \int_I \varphi(y)\, P(s,x,dy) \quad \text{for } \varphi \in \mathbf{B}(I). \tag{1}$$

This is the conditional expectation of $\varphi(X_s)$ given $X_0 = x$, i.e., $\mathcal{T}_s\varphi(x) = E[\varphi(X_s)\,|X_0 = x]$. Note that the conditional expectation itself may be defined for some unbounded function $\varphi \notin \mathbf{B}(I)$ (as long as φ is integrable with respect to the transition function). However, to characterize Markov processes based on the operators, it is sufficient to consider a space of bounded functions. By the law of iterated expectations (or by the third property of the transition function in Assumption 1), $\{\mathcal{T}_s\}(:= \{\mathcal{T}_s\}_{s\geq 0})$ satisfies a *semigroup* property, i.e., $\mathcal{T}_{s+t} = \mathcal{T}_s\mathcal{T}_t = \mathcal{T}_t\mathcal{T}_s$, for any $s, t (\in [0,\infty))$. We call $\{\mathcal{T}_s\}$ *a semigroup of the conditional expectations* associated with the Markov process $\{X_s\}$, or simply *a semigroup*. For the semigroup, we define its *infinitesimal generator* $\mathcal{A} : \mathfrak{D}(\mathcal{A}) (\subset \mathbf{B}(I)) \to \mathbf{B}(I)$ by

$$\mathcal{A}\varphi(x) := \lim_{\Delta\to 0+} \frac{\mathcal{T}_\Delta \varphi(x) - \varphi(x)}{\Delta} = \lim_{\Delta\to 0+} \frac{E[\varphi(X_\Delta)\,|X_0 = x] - \varphi(x)}{\Delta}, \tag{2}$$

where $\mathfrak{D}(\mathcal{A})$ is the domain of $\mathcal{A}$, i.e., a subset of $\mathbf{B}(I)$ for which the convergence on the right-hand side (RHS) of (2) takes place with respect to the sup-norm $||\cdot||$. We call an element φ of $\mathfrak{D}(\mathcal{A})$ as a *test* function. A semigroup $\{\mathcal{T}_s\}$ is called a *Feller* semigroup when it satisfies the following conditions: (i) $\mathcal{T}_s : \hat{\mathbf{C}}(I) \to \hat{\mathbf{C}}(I)$ for each s; (ii) for $\varphi \in \hat{\mathbf{C}}(I)$, $||\mathcal{T}_s\varphi - \varphi|| \to 0$ as $s \to 0$, where $\hat{\mathbf{C}}(I) (\subset \mathbf{B}(I))$ is the space of continuous functions vanishing at infinity (i.e., $\lim_{|x|\to\infty} |f(x)| = 0$) with the sup-norm $||f|| := \sup_{x\in I} |f(x)|$. A process is also called a Feller process if its associated conditional expectation operator satisfies (i) and (ii). Note that for any Feller semigroup $\{\mathcal{T}_s\}$ and $\varphi \in \hat{\mathbf{C}}(I)$, $\mathcal{A}\varphi$ is necessarily in $\hat{\mathbf{C}}(I)$ if it is well-defined, since $[\mathcal{T}_\Delta\varphi(x) - \varphi(x)]/\Delta$ is continuous and vanishing at infinity for each Δ (by the Feller property (i)) and $\mathcal{A}\varphi(x)$ is its uniform limit. We also note that by time-homogeneity, it holds that $\mathcal{A}\varphi(x) = \lim_{\Delta\to 0+} [\mathcal{T}_{s+\Delta}\varphi(x) - \mathcal{T}_s\varphi(x)]/\Delta$ for any $s > 0$.

[4] That is, $\Pr[\zeta = \infty] = 1$, where $\zeta := \inf\{s \in [0,\infty) : X_s \notin I \text{ or } \liminf_{u\to s} X_u \notin I\}$ is the *lifetime* of the process.

We have defined the generator $\mathcal{A}$ on the space $\mathbf{B}(I)$, but it is often sufficient to look at its restriction on some subspace of $\mathbf{B}(I)$ to characterize Markov processes (see, e.g., discussions in Sec. 5 of Dynkin, 1956 or in Ch. II of Dynkin, 1965). In the sequel, we only consider $\mathcal{A}$ on $\hat{\mathbf{C}}(I)$, regarding $\mathcal{A}$ as a mapping on $\hat{\mathbf{C}}(I)$ to $\hat{\mathbf{C}}(I)$. This restriction can be justified since we hereafter pay attention to time-homogeneous Markov processes which are in the class of Feller processes.[5] Note that the infinitesimal generators on $\hat{\mathbf{C}}(I)$ and the corresponding domains can fully characterize the class of Feller processes in the following sense: if the infinitesimal generators of two processes are equal (with the same domain), then the transition functions are also equal (see, e.g., Sec. 5 of Ch. II of Dynkin, 1965), implying a one-to-one mapping between infinitesimal generators and Feller processes.[6] That is, the knowledge on the form of $\mathcal{A}$ and $\mathfrak{D}(\mathcal{A})$ allows us to recover the complete form of the transition function. Our restriction on the Feller class is not strong, and it will not rule out any interesting Markov processes. Many of Markov processes used in the economics/finance literature are actually Feller. In particular, many processes represented by stochastic differential equations (SDEs) of the diffusion type or general Lévy type turn out to be Feller under weak conditions (see, e.g., V. 22 of Rogers and Williams, 2000, and Ch. 6 of Applebaum, 2009). Even when a process defined by some SDE does not satisfy the Feller-semigroup properties (i) and (ii), we may be able to construct another (modified) process whose behaviors are very close to those of the original process, so that they are almost indistinguishable from an empirical/statistical point of view. For example, if coefficients of the original SDE possess a sort of continuity property, this may be achieved by the method of dapmping, as poposed in Li (2010). On the other hand, the Feller restriction ensures the continuity of $\mathcal{A}\varphi$ under the sup-norm based definition of $\mathcal{A}$ in (2).[7] This continuity property is useful for avoiding some technical difficulties and allows us to develop identification and asymptotic results more easily in the subsequent sections.

Before concluding this section, we provide some discussions on the form of the infinitesimal generator (and its domain). It is not easy to know the precise forms of $\mathcal{A}$ and $\mathfrak{D}(\mathcal{A})$ for a general Feller process. However, if $\mathfrak{D}(\mathcal{A})$ contains $\mathbf{C}_K^\infty(\mathbb{R})$ (the set of infinitely continuously differentiable functions with compact support), where we set $I = \mathbb{R}$, the restriction of $\mathcal{A}$ on $\mathbf{C}_K^\infty(\mathbb{R})$ is known to take the following form

[5] Note the following facts: i) any Feller process has a modification whose sample path is càdlàg (right continuous with left limits); ii) any Feller process whose path is càdlàg is also a (strong) Markov process. For these results, see, e.g., Theorems 19.15 and 19.17 of Kallenberg (2002), with noting that the lifetime $\zeta = \infty$ almost surely in our case. By identifying a Feller process as its càdlàg modification, we can say Feller processes represent a subclass of (strong) Markov processes.

[6] For a general class of Markov processes, we only have the weaker assertion that the infinitesimal generator (and its domain) may determine the finite-dimensional distributions of the process. For details of this point, refer to arguments on the Hill-Yoshida theorem and transitions functions in Ethier and Kurtz (1986) (Sec. 2 of Ch. 1 and Sec. 1 of Ch. 4, respectively).

[7] Note that there exist some other definitions of infinitesimal generators. For example, we can use the $L_2(Q)$ norm to define the convergence in (2), instead of the sup-norm (Q is the invariant measure of the process, the existence of which is supposed). However, the limit $\mathcal{A}\varphi$ is not necessarily continuous under this definition.

(under Assumption 1): for any $\varphi(\cdot) \in \mathbf{C}_K^\infty(\mathbb{R})$,

$$\begin{aligned}\mathcal{A}\varphi(x) = L\varphi(x) &:= \alpha(x)\varphi'(x) + (1/2)\beta(x)\varphi''(x) \\ &+ \int_{\mathbb{R}\setminus\{0\}} \left[\varphi(x+z) - \varphi(x) - \mathbf{1}_{\{|z|\le 1\}} z\varphi'(x)\right] l(x, dz), \end{aligned} \tag{3}$$

where $\alpha(\cdot)$ is a continuous function; $\beta(\cdot)$ is a continuous and non-negative function; $\mathbf{1}_{\{|z|\le 1\}}$ is the indicator function ($= 1$ if $|z| \le 1$, and $= 0$ otherwise); and $l(\cdot,\cdot)$ is a Lévy kernel ($l : \mathbb{R} \times \mathfrak{B}(\mathbb{R}\setminus\{0\}) \to \mathbb{R}^+$), i.e., $l(x,\cdot)$ is a Borel measure on $\mathbb{R}\setminus\{0\}$ with satisfying

$$\int_{\mathbb{R}\setminus\{0\}} \left[1 \wedge z^2\right] l(x, dz) < \infty, \tag{4}$$

for each $x \in \mathbb{R}$.[8] The integro-differential operator L in the form of (3) is said to be of a Lévy type. This representation result for $\mathcal{A}$ follows from the Courrège theorem (see Sec. 3.5 of Applebaum, 2009 and Sec. 4.5 of Jacob, 2001) and the fact that $\mathcal{A}\varphi$ is continuous. We may be able to interpret the Lévy kernel $l(x, dz)$ as representing the expected number of jumps (conditional on the current state x) the size of which is in the (small) interval "dz" per unit of time. The only restriction on ν is (4), and it allows for the case with $\int_{\mathbb{R}\setminus\{0\}} l(x, dz) = \infty$, which corresponds to an infinite number of jumps within a finite time interval. Note that the requirement that $\mathbf{C}_K^\infty(I) \subset \mathfrak{D}(\mathcal{A})$ is weak and most Feller processes known in the literature should satisfy it. Indeed, this assumption seems to be used as a base for developing various theories on Feller processes (see, e.g., Stroock, 1975; Taira, 1992; Böttcher and Schnurr, 2010).[9]

The general form of (3) includes several special cases. An important one is a generator of the diffusion type. We say that the generator is of the diffusion type, if there exist some continous function $\mu(\cdot)$ and some non-negative continuous function $\sigma^2(\cdot)$ such that for any $\varphi(\cdot) \in \mathbf{C}_K^\infty(\mathbb{R})(\subset \mathfrak{D}(\mathcal{A}))$,

$$\mathcal{A}\varphi(x) = G\varphi(x) := \mu(x)\varphi'(x) + \sigma^2(x)\varphi''(x)/2. \tag{5}$$

In this case, $\{X_s\}$ is called a diffusion process, and any of its realized path is continuous on $[0,\infty)$ almost surely. Conversely, if any path of a Feller process $\{X_s\}$ (satisfying Assumption 1) is continuous on $[0,\infty)$, we can also say that there exist some continuous functions μ and σ^2 with σ^2 non-negative such that $\mathcal{A}\varphi(x) = G\varphi(x)$ for $\varphi(\cdot) \in \mathbf{C}_K^\infty(\mathbb{R})$.[10] For some diffusion processes, we may be able to know the precise forms of $\mathcal{A}$ and $\mathfrak{D}(\mathcal{A})$ under several (boundary) conditions on μ and σ^2 (see, e.g., Sec. 1-2 of Ch. 8 of Etheir and Kurtz, 1986)

[8] Instead of the truncation function $\mathbf{1}_{\{|z|\le 1\}}$ in (3), we may be able to use some other function (say, some smoothed version of it, or $1/\left[1+z^2\right]$) with some minor modification of $\alpha(z)$. By the integrability condition (4), such modification is possible. Note also that (4) can be (equivalently) written as $\int_{\mathbb{R}\setminus\{0\}} z^2\left[1+z^2\right]^{-1} v(x, dz) < \infty$ (see Sec. 1.2.4 of Applebaum, 2009).

[9] Recall that (i) of Assumption 1 implies the no-killing condition (i.e., the lifetime of the process ζ is infinite). Without this condition, we generally need an additional component $c(x)\varphi(x)$ in the RHS of (3), where $c(\cdot)$ is some continuous function with $c(x) \le 0$.

[10] For these statements, see Theorems 13.3 and 13.5 in Ch. III of Rogers and Williams (2000).

We can also think of a class of pure jump processes. For example, let $\{X_s\}$ be a Markov jump process described by two components q $(: I \times \mathfrak{B}(I) \to \mathbb{R})$ and $\lambda(: I \to \mathbb{R})$, where $q(x, \cdot)$ is a probability measure on $\mathfrak{B}(I)$ for each $x \in I$ and $\lambda(\cdot)$ is a bounded continuous function. A Poisson process with intensity parameter $\lambda(x)$ (when the current state is $x \in I$) determines the timing of jump changes. If a jump occurs, then the transition probability from the state x to Γ is given by $q(x, \Gamma)$. The infinitesimal generator of this process has the following form:

$$\mathcal{A}\varphi(x) = \lambda(x) \int_{\mathbb{R}} [\varphi(y) - \varphi(x)] \, q(x, dy), \tag{6}$$

which is also a special case of (3) (upon suitable reparametrization). Ethier and Kurtz (1986) provid more details on this type of process in Sec. 2 of Ch. 4 and Sec. 3 of Ch. 8, where we can find the full characterization of $\mathfrak{D}(\mathcal{A})$.

3 Identifying the stationarity property

To identify the stationary property of the process $\{X_s\}$, we use the infinitesimal generator introduced in the previous section. Now, to formally state our null and alternative hypotheses, we also set out the following condition:

Assumption 2 *(i)* $\{X_s\}$ *is (Harris) recurrent with its invariant* σ*-finite measure* Π *on* $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$*, i.e., for any* $\Gamma \in \mathfrak{B}(\mathbb{R})$ *with* $\Pi(\Gamma) > 0$,

$$\Pr\{X_s \in \Gamma \text{ infinitely often}\} = 1.$$

(ii) The invariant measure Π *of* $\{X_s\}$ *has the density function* π *which is continuous and uniformly bounded over* $\mathbb{R}$*, i.e.,* $\Pi(\Gamma) = \int_\Gamma \pi(x)\,dx$ *for any* $\Gamma \in \mathfrak{B}(\mathbb{R})$.

(i) of Assumption 2 may be called Π-irreducible in the Markov chain terminology, and it is interpreted as $\{X_s\}$ (re-)visits any arbitrary set in the state space $\mathbb{R}$ within some finite time and infinitely many times over the time span $[0, \infty)$. If the process is not recurrent, it is called *transient*, i.e., the process does not necessarily revisit every set in the state space, tending to ∞ or $-\infty$ (in our case where $I = \mathbb{R}$). The condition (i) also implies that no absorption occurs at any point x and the process does not forever remain at the same point, i.e., for any $x \in I$, there exists some $s \in [0, \infty)$ such that $P(s, x, \{x\}) < 1$. The measure Π is said to be *invariant* when it satisfies $\Pi(\Gamma) = \int_{\mathbb{R}} P(s, x, \Gamma)\,\Pi(dx)$ for any $\Gamma \in \mathfrak{B}(\mathbb{R})$, where Π is unique up to constant multiples (see, e.g., Sec. 1 of Höpfner and Löcherbach, 2003). A recurrent process is called *positive recurrent* (or *ergodic*) if $\Pi(\mathbb{R}) < \infty$ and *null recurrent* if $\Pi(\mathbb{R}) = \infty$. When $\{X_s\}$ is positive recurrent, Π may be interpreted as the invariant probability measure of the process (upon suitable normalization). In this case, we therefore regard π as the probability density.

When $\{X_s\}$ is obtained as a solution to some SDE of the diffusion type, a necessary and sufficient condition (in terms of coefficients of the SDE) for the process to be recurrent is well-known (see Sec.

5.5 of Karatzas and Shreve, 1991). If $\{X_s\}$ is a solution to a SDE of the (so-called) jump-diffusion type, Wee (1999, 2000) provides some sufficient conditions for the recurrency.

For a class of Feller processes with satisfying Assumptions 1 and 2, we consider the following null and alternative hypotheses:

The null hypothesis H_0: $\{X_s\}$ is a strictly stationary process, i.e., the probability

$$\Pr\left[X_{t+s_1} \in \Gamma_1, X_{t+s_2} \in \Gamma_2, \ldots, X_{t+s_k} \in \Gamma_k\right] \tag{7}$$

is independent of $t \geq 0$ for any $k\,(=1,2,\ldots)$, $0 \leq s_1 < \cdots < s_k < \infty$, and $\Gamma_1, \ldots \Gamma_k \in \mathfrak{B}(I)$.

The alternative hypothesis H_1: $\{X_s\}$ is null recurrent.

Our definition of the strict stationarity in (7) is standard. Under the hypothesis H_0, the invariant density π is the same as the marginal probability density of X_0 and therefore, it must be integrable, i.e., $\int_{\mathbb{R}} \pi(x)\, dx = \Pi(\mathbb{R}) < \infty$. A simple example in the alternative class is a Brownian motion (with no drift), whose invariant density $\pi(x) = c$ for any $x \in \mathbb{R}$ with some constant $c > 0$ (and therefore $\int_{\mathbb{R}} \pi(x)\, dx = \infty$). This process is a continuous-time counterpart of a unit-root process. In the literature on discrete time series econometrics, unit-root processes (and some of their relatives) are often referred to as *stochastic trends*. Our alternative class may be interpreted as the class of continuous-time counterparts of such stochastic trends.

Note that Assumption 2 excludes processes with obvious upward or downward trends. For example, the geometric Brownian motion $dX_s = \alpha X_s ds + \beta X_s dW_s$ is excluded unless $\alpha - \beta^2/2 = 0$.[11] However, we impose Assumption 2 to clarify the class of processes against which our test is consistent and to develop some sensible distribution theory. Our test may have some power for a certain class of nonstationary processes. Indeed, we can show that the test can reject the geometric Brownian motion. Note also that the strict stationarity is imposed to develop our identification theorem as below. We can show that our proposed test has no asymptotic power against processes which are stationary only asymptotically (but not strictly stationary) under some condition.[12] Processes that are stationary in the strict or asymptotic sense may be said to represent the class of processes with stability. In this respect, our test can be interpreted as a test for examining the stability of $\{X_s\}$.

Our identification of the stationarity property of $\{X_s\}$ is based on the following result:

[11] If $\alpha - \beta^2/2 = 0$, the geometric Brownian motion is null recurrent, but otherwise, it has a diverging trend to ∞ or $-\infty$ ($\{W_s\}$ is a Brownian motion, and $\alpha \in \mathbb{R}$ and $\beta > 0$).

[12] For example, we can think of a process that is ergodic but not initialized by the invariant distribution. By using the so-called strong Doeblin condition (as in Kristensen, 2009), we will be able to verify that the test has no asymptotic power against this process.

Lemma 1 *Let $\{X_s\}$ be a continuous-time Feller process with the corresponding infinitesimal generator $\mathcal{A}$ and its domain $\mathfrak{D}(\mathcal{A})$. Suppose that $\{X_s\}$ satisfies Assumptions 1 and 2. Then, $\{X_s\}$ is strictly stationary with the invariant (probability) density π if and only if*

$$\int_I \mathcal{A}\varphi(x)\,\pi(x)\,dx \;\; (= E[\mathcal{A}\varphi(X_s)]) = 0, \tag{8}$$

for every test function φ in $\mathfrak{D}(\mathcal{A})$.

This lemma is a version of Proposition 9.2 in Ethier and Kurtz (1986, Ch. 4), and we omit the proof for brevity.[13] Our testing procedure is to nonparametrically estimate $\int \mathcal{A}\varphi(x)\,\pi(x)\,dx$ and then check whether its estimate is close to zero. Apparently, from the "if and only if" statement of Lemma 1, we need to consider various test functions and compute corresponding unconditional moments to check the stationarity of $\{X_s\}$. However, it is not easy to check the equality (8) for all test functions in $\mathfrak{D}(\mathcal{A})$. The domain of $\mathcal{A}$ generally consists of an infinite number of functions in $\hat{\mathbf{C}}(I)$. In first place, it is difficult to know the precise form of $\mathfrak{D}(\mathcal{A})$. In some limited cases, where $\mathcal{A}$ is known to take a certain convenient form, we might be able to obtain the full characterization of of $\mathcal{A}$ (see, e.g., Ch. 8 of Ethier and Kurtz, 1986). However, even in such limited cases, we generally need several (so-called) boundary/lateral conditions to characterize $\mathfrak{D}(\mathcal{A})$, and these conditions often take quite intricate forms and are not necessarily convenient for our purpose to develop a statistical testing procedure.[14]

We note that Hansen and Scheinkmann (1995) proposed to use the restriction (8) to construct moment conditions for estimating parametric stationary Markov processes. For identifying a parametric model, we do not necessarily examine all the test functions in $\mathfrak{D}(\mathcal{A})$. It is often enough to look at only some finite number of test functions. However, since our problem is to check the stationarity property, which is indeed a nonparametric restriction, we need to consider infinite many number of test functions. However, again it is not easy to look at such many functions.

One way to overcome this difficulty is to use some smaller set of test functions, but such a reduction may result in losing information and yielding lower power of the corresponding test. Fortunately, we can construct a reduced class of test functions without any information loss. Our approach is based on a result from *approximation theory.* The next lemma states that any k-times continuously differentiable function φ in $\hat{\mathbf{C}}(\mathbb{R})$ can be well approximated by a sequence of weighted polynomial functions. Let

$$\mathfrak{w}(x) := \exp\left\{-x^2/2\right\}/\sqrt{2\pi}, \tag{9}$$

[13] Note that in Ethier and Kurtz (1986), the result requires that the martingale problem associated to the generator $\mathcal{A}$ is *well posed*, i.e., there exists some solution to the martingale problem $\{X_s\}$, and any other solution has the same finite-dimensional distribution as $\{X_s\}$ (for details on the martingale problem, see Ch. 4 of Ethier and Kurtz, 1986). The well-posedness is imposed because they start with a generator $\mathcal{A}$, and construct $\{X_s\}$ as a solution to the martingale problem. In general, it is not easy to check the well-posedness of some given generator. On the other hand, in this paper, we start with some Feller process $\{X_s\}$ (defined through the transition probability) on the filtered probability space $(\Omega, \mathfrak{F}, \{\mathfrak{F}_s\}, \Pr)$ and therefore, we do not need to consider the well-posedness in the martingale problem.

[14] For a general form of the infinitesimal generator of a Feller process on some subset of $\hat{\mathbf{C}}(I)$, see, e.g., Theorem 1.13 in Ch. VII of Revus and Yor (1999). For general boundary conditions, see, e.g., Taira (1992) and references therein.

which is the density of the standard normal. Using $\mathfrak{w}$ as a weighting function, we obtain the following result:

Lemma 2 *Let φ be an arbitrary function (in $\hat{\mathbf{C}}(\mathbb{R})$) which is k-times continuously differentiable ($k \geq 0$). Then, for each φ, there exists a sequence of functions, $\{L_J(\cdot) : J = k+1, k+2, \dots\}$, such that each $L_J(\cdot)$ is a polynomial function of the degree at most $J-1$, and*

$$\sum\nolimits_{i=0}^{k} \sup\nolimits_{x\in\mathbb{R}} \left|\varphi^{(i)}(x) - H_J^{(i)}(x)\right| \to 0 \quad as \quad J \to \infty, \tag{10}$$

where $H_J(x) := L_J(x)\,\mathfrak{w}(x)$, and $\varphi^{(i)}$ and $H_J^{(i)}$ are the i-th order derivatives of φ and H_J, respectively.

The proof is provided in the Appendix. The result that some appropriate polynomial function well approximates a certain sort of smooth function is widely known. The lemma strengthens this result by providing the simultaneous approximation of the function itself and its derivatives. A key assumption for the result is that $\varphi(x)$ vanishes as $|x| \to \infty$, and the weighting function $\mathfrak{w}$ plays a role in controlling aberrant behaviors of polynomial functions in the tail region.

The result of Lemma 2 suggests that it is sufficient to look at the set of weighted polynomial functions (instead of the whole set $\mathfrak{D}(\mathcal{A}) \subset \hat{\mathbf{C}}(\mathbb{R})$) in order to check the stationarity. This idea is indeed correct, but the set of weighted polynomial functions is still large and may not be tractable enough. Therefore, we consider a further reduction. Let $\{\phi(\cdot;\theta) : \theta \in \Theta\}$ be a set of functions indexed by ξ such that

$$\phi(x;\theta) := \exp\{\theta x\}\,\mathfrak{w}(x) = \exp\left\{\theta x - x^2/2\right\}/\sqrt{2\pi}, \tag{11}$$

with Θ being some bounded interval on $\mathbb{R}$. Recall that the exponential function may be expressed as an infinite series of polynomial functions. This fact and the result of Lemma 2 allow us to develop a convenient theorem to check the stationarity property of the process under the following conditions:

Assumption 3 *Let $\mathcal{A}$ and $\mathfrak{D}(\mathcal{A})$ be respectively the infinitesimal generator and its domain of a Feller process $\{X_s\}$. (i) For any $\theta \in \Theta$, $\phi(\cdot;\theta) \in \mathfrak{D}(\mathcal{A})$, and for any non-negative integer $l (\geq 0)$, $g_l(\cdot) \in \mathfrak{D}(\mathcal{A})$ and $\left|\int_{\mathbb{R}} \mathcal{A}g_l(x)\,\pi(x)\,dx\right| < \infty$, where $g_l(x) := x^l \mathfrak{w}(x)$. (ii) Let $\varphi(\cdot)$ be an (arbitrary) element of $\mathfrak{D}(\mathcal{A})$ which is k-times continuously differentiable with some $k \geq 0$. If there is a sequence of functions $\{\varphi_J(\cdot)\}$ approximating $\varphi(\cdot)$ such that each $\varphi_J(\cdot) \in \mathfrak{D}(\mathcal{A})$, and*

$$\sum\nolimits_{i=0}^{k} \sup\nolimits_{x\in\mathbb{R}} \left|\varphi^{(i)}(x) - \varphi_J^{(i)}(x)\right| \to 0 \quad as \quad J \to \infty, \tag{12}$$

then, it holds that

$$\sup\nolimits_{x\in\mathbb{R}} \left|\mathcal{A}\varphi(x) - \mathcal{A}\varphi_J(x)\right| \to 0 \quad as \quad J \to \infty. \tag{13}$$

(i) of Assumption 3 is fairly weak and should be satisfied by many Feller process. In particular, we note that $\phi(x;\theta)$ and $g_l(x)$ are infinitely differentiable and possess exponential decay rates (as $x \to \infty$). Such functions are in the domain of the generator in all examples of Feller processes (with state space

$\mathbb{R}$) in Ch. 8 of Ethier and Kurtz (1986). Indeed, the author does know of an example that would violate the condition (i). (ii) of Assumption 3 is also not restrictive in view of the general form of $\mathcal{A}$ given in (3). To see this point, suppose that the generator of $\{X_s\}$ is given as L in (3) for *any* test function $\varphi \in \mathfrak{D}(\mathcal{A})$. In this case, φ is at least twice continuously differentiable ($k = 2$), and we can check (13) since $L\varphi, \varphi_J(x) \in \hat{\mathbf{C}}(\mathbb{R})$.

Given Lemma 2 and Assumption 3, we can now state our identification theorem:

Theorem 1 *Let $\{X_s\}$ be a Feller process satisfying the conditions in Assumptions 1 and 2 with the infinitesimal generator $\mathcal{A}$ and its domain $\mathfrak{D}(\mathcal{A})$. Suppose that $\mathcal{A}$ and $\mathfrak{D}(\mathcal{A})$ satisfy the conditions in Assumption 3. Let Θ be any (arbitrary) finite interval on $\mathbb{R}$ which contains a neighborhood of 0. Then, it holds that*

$$\int_{\mathbb{R}} \mathcal{A}\varphi(x)\,\pi(x)\,dx \neq 0,$$

for some test function $\varphi \in \mathfrak{D}(\mathcal{A})$, if and only if there exists some $\bar{\theta} > 0$ ($\bar{\theta}$ may be arbitrarily close to zero) and for any $\theta \in S(\bar{\theta}) := [-\bar{\theta}, \bar{\theta}]$,

$$\int_{\mathbb{R}} \mathcal{A}\phi(x;\theta)\,\pi(x)\,dx \neq 0.$$

The proof of the theorem is provided in the Appendix. An intuition behind this result is that any function in $\mathfrak{D}(\mathcal{A}) \subset \hat{\mathbf{C}}(\mathbb{R})$ has a component *correlated* with a parametric function family $\{\phi(\cdot;\theta)\}$ in a certain sense. The result of this theorem allows us to construct a feasible but consistent testing procedure. The set of functions we need to check is effectively reduced to $\{\phi(\cdot;\theta)\}$, a set of parameterized functions, while the "if and only if" statement still holds.

Since $\int_{\mathbb{R}} \mathcal{A}\phi(x;\theta)\,\pi(x)\,dx = 0$ for any $\theta \in \Theta$ under the null hypothesis, θ may be called a nuisance parameter. A similar technique can be found in the so-called Bierens approach (or the nuisance parameter approach), named based on Bierens' (1982, 1990) seminal work (see also Andrews and Ploberger, 1994; Bierens and Ginther, 2001; Bierens and Ploberger, 1997; Boning and Sowell, 1999; Chen and Fan, 1999; De Jong, 1996; De Jong and Bierens, 1994; Hansen, 1996; Kasparis, 2010; Stinchcombe and White, 1998). These papers consider testing procedures to examine parametric specifications of conditional moment functions (or regression functions). While the result of Theorem 1 is (at least apparently) similar to the results of the Bierens approach, it is not an obvious extension since we work with more complicated functional operators (differential operators defined via conditional moment functions) instead of conditional moment functions themselves. This complication requires us to use approximation theory as in Lemma 2, but not unnecessary in the Bierens approach.

Our test function defined in (11) is the product of the (rescaled) exponential function $\exp\{\theta x\}$ and the weighting function $\mathfrak{w}(x)$. As in the case of the Bierens approach, some suitable function, such as $\cos(\theta x) + \sin(\theta x)$, $1/[1 + \exp\{c - \theta x\}]$ ($c \neq 0$), may replace the exponential function. As shown in Theorem 3.1 of Stinchcombe and White (1998), any function may be used in the Bierens approach as long as the linear span of indexed functions is dense in the weak topology (in the space of bounded functions satisfying some sort of measurability). In our case, however, some function that is allowed in the Bierens approach may not be used. For example, obviously, because we are considering the differential operator, we cannot use a function $\mathbf{1}\{x \leq \theta\}$. At least from the viewpoint of our proofs, it seems necessarily to work with a class of functions that admit a sort of approximation result as in Lemma 2. The weighting function $\mathfrak{w}(x)$ in (9) is chosen only due to its familiarity in the statistical literature. We can choose some other type of function, say, a *Freud* type weight (see Balazs, 2004; Szabados, 1997). Such a choice will also allow us to prove an approximation result as in Lemma 2 and develop its corresponding identification result for the stationarity. While it is interesting to investigate what type of test function may be used in our context, this would require more extra work and we leave it to future work.

4 A test statistic and its asymptotic behavior

4.1 A test statistic

Motivated by the identification result in the previous section, in this section, we construct a test statistic to examine the stationary property of the process. For this purpose, we consider in particular the following quantity:

$$\int_{\Theta}\left\{\int_{\mathbb{R}}\mathcal{A}\phi(x;\theta)\,\pi(x)\,dx\right\}^2 d\theta.$$

By Theorem 1, this quantity is zero if and only if the null hypothesis H_0 is true. We construct an emprical counterpart of this and use its normalized version as our test statistic. In doing so, we suppose that the process is discretely sampled and we can obtain $\{X_{i\Delta}\}_{i=0}^{n}$, where $(n+1)$ is the number of observations and Δ is the observation interval. The observation time span is represented by $T(:= n\Delta)$. Given $\{X_{i\Delta}\}$, we estimate

$$\Psi(x;\theta) := \mathcal{A}\phi(x;\theta)\times\pi(x),$$

by the following kernel-based estimator:

$$\hat{\Psi}(x;\theta) := T^{-1}\sum\nolimits_{i=0}^{n-1}K_h(X_{i\Delta}-x)\left[\rho(X_{i\Delta})/\rho(x)\right]\left[\phi\left(X_{(i+1)\Delta};\theta\right)-\phi(X_{i\Delta};\theta)\right],$$

where $K_h(z) := K(z/h)/h$; K is a kernel function; and h is a bandwidth (smoothing parameter); and $\rho(\cdot)$ is a weighting function, which is not a constant, with $\rho(x) > 0$ for any x. This $\hat{\Psi}(x;\theta)$ is a consistent estimator of $\Psi(x;\theta)$, and therefore, under the null hypothesis, its integral

$$\int_{\Theta}\left\{\int_{\mathbb{R}}\hat{\Psi}(x;\theta)\,dx\right\}^2 d\theta \tag{14}$$

is expected to approximate $\int_{\Theta}\{E[\mathcal{A}\phi(X_s;\theta)]\}^2\,d\theta = 0$, where the integrals with respect to x and θ are computed by using some numerical method. Note that we have introduced the weighting function ρ to utilize the variability of increments of the process. We might be able to consider an estimator without $\rho\left(X_{i\Delta}\right)/\rho\left(x\right)$, such as

$$\tilde{\Psi}\left(x;\theta\right) := T^{-1}\sum\nolimits_{i=0}^{n-1} K_h\left(X_{i\Delta}-x\right)\left[\phi\left(X_{(i+1)\Delta};\theta\right)-\phi\left(X_{i\Delta};\theta\right)\right].$$

However, since $\int_{\mathbb{R}} K_h\left(X_{i\Delta}-x\right)dx = 1$ for any $X_{i\Delta}$, which follows from the convolution property and the condition that $\int K\left(z\right)dz = 1$, the integral of this estimator with respect to x is simply reduced to

$$\int_{\mathbb{R}} \tilde{\Psi}\left(x;\theta\right)dx = T^{-1}\left[\phi\left(X_{n\Delta};\theta\right)-\phi\left(X_0;\theta\right)\right]. \tag{15}$$

This quantity does not seem to exploit enough information from the data, relying only on the first and end observations. Note that by using a non-constant ρ, we can let a test based on $\hat{\Psi}\left(x;\theta\right)$ have a power property for some class of cyclic/periodic processes. For example, consider the case where $X_{i\Delta} = \sin\left(i\pi/2\right)$. Then, it holds that $\int \tilde{\Psi}\left(x;\theta\right)dx \neq 0$ for odd n, but $= 0$ for even n, implying no consistency against this cyclic process. For a strictly monotone function ρ, we can show that the test based on $\hat{\Psi}\left(x;\theta\right)$ is consistent against this $X_{i\Delta} = \sin\left(i\pi/2\right)$. We conjecture that by setting the weight function as a strictly monotone function, our test will have consistency against some class of processes which include sorts of cyclical/periodic components as $\sin\left(i\pi/2\right)$. That is, the test will be consistent against not only the class of processes specified by Assumption 2 and the alternative condition H_1, but also some other class of processes (note that the cyclic process $\sin\left(i\pi/2\right)$ is deterministic and does not satisfy Assumption 2).

To develop a formal statistical testing procedure, we investigate the asymptotic behavior of (14) by considering its scaled version as our test statistic:

$$\hat{J} := \int_{\Theta}\left\{\sqrt{T}\int_{\mathbb{R}}\hat{\Psi}\left(x;\theta\right)dx\right\}^2 d\theta \bigg/ \int_{\Theta}\hat{\lambda}\left(\theta\right)d\theta, \tag{16}$$

where the scaling factor in the denominator is defined as

$$\hat{\lambda}\left(\theta\right) := T^{-1}\sum\nolimits_{i=0}^{n-1}\left|\int_{\mathbb{R}} K_h\left(X_{i\Delta}-x\right)\rho^{-1}\left(x\right)dx\rho\left(X_{i\Delta}\right)\left[\phi\left(X_{(i+1)\Delta};\theta\right)-\phi\left(X_{i\Delta};\theta\right)\right]\right|^2.$$

4.2 The SDE representation of a Feller process

To investigate the asymptotic behavior of the statistic $\hat{J}$, we use the fact that any Feller process may be represented by some sort of SDE. Given the SDE representation as below, we can use the Ito formula, as well as some limit results for additive functionals of the Markov process (as developed in Höpfner and Löcherbach, 2003).

Lemma 3 *Let $\{X_s\}$ be a Feller process satisfying Assumptions 1 and (2), and let $\mathcal{A}$ and $\mathfrak{D}\left(\mathcal{A}\right)$ be the infinitesimal generator of $\{X_s\}$ and its corresponding domain. Suppose that $\mathbf{C}_K^{\infty}\left(\mathbb{R}\right) \subset \mathfrak{D}\left(\mathcal{A}\right)$. Then,*

$\{X_s\}$ satisfies some SDE of the following type:

$$dX_s = \mu (X_{s-}) \, ds + \sigma (X_{s-}) \, dW_s$$
$$+ \int_{|\gamma(X_{s-},z)| \in (0,1]} \gamma (X_{s-}, z) \, \tilde{N} (ds, dz) + \int_{|\gamma(X_{s-},z)|>1} \gamma (X_{s-}, z) \, N (ds, dz) , \qquad (17)$$

where $\{X_{s-}\}$ is a càglàd version of $\{X_s\}$ (a càglàd process is a process whose path is left-continuous with right limits almost surely); $\mu(\cdot)$ and $\sigma(\cdot)$ are continuous functions with $\sigma(x) \geq 0$ for any $x \in \mathbb{R}$; $\{W_s\}_{s \geq 0} := \{W_s\}$ is a Brownian motion; $\tilde{N}(\cdot, \cdot)$ is the compensated version of a Poisson random measure $N(\cdot, \cdot)$ on $\mathbb{R}^+ \times \mathbb{R} \backslash \{0\}$ which is independent of $\{W_s\}$ and whose intensity measure is $\nu(dx)\, ds = E[N(ds, dx)]$ (i.e., $\tilde{N}(ds, dz) = N(ds, dz) - \nu(dz)\, ds$, and $\nu(A) = E[N(1, A)]$ for any Borel set $A \in \mathfrak{B}(\mathbb{R} \backslash \{0\})$); v is a sigma finite measure on $\mathbb{R} \backslash \{0\}$; and $\gamma(\cdot, \cdot)$ is a measurable function ($\mathbb{R}^2 \to \mathbb{R}$).

The proof of the lemma is provided in the Appendix. Note that the result here only asserts that a Feller process $\{X_s\}$ is a weak solution to (17). It does not claim its uniqueness in either the weak or the strong sense.[15] To achieve the existence and uniqueness of a solution to (17), we generally need to impose some conditions on the growth rate and/or the smoothness of the functions (μ, σ^2, and γ) and the measure v.[16] However, we do not pursue such conditions in this paper. While the existence of a unique strong solution to (17) (for some Brownian motion $\{W_s\}$ and some Poisson random measure given) is required in some specific applications, that of a weak solution is often sufficient for many econometric/statistical purposes. This is also the case here. We only require that the process have a representation by some SDE of the type (17). When deriving distributional theory of our test statistic, we use this SDE expression and the Ito formula (note that the Ito formula's prerequisite is not relevant to the uniqueness of SDE solutions, see, e.g., Ch. 4 of Applebaum, 2009).

For $\{X_s\}$ to possess a SDE-based expression of the type (17), the no-killing and no-absorption conditions (implied by Assumptions 1 and 2) play an important role. If either/both killing or absorption may happen, the process is not generally expressed by (17). Setting $I = \mathbb{R}$ also makes our arguments easier. If I has a finite endpoint such as $(0, \infty)$, $[0, \infty)$, and $[0, 1]$, the SDE representation of $\{X_s\}$ may require some additional (local time based) component.[17] On the other hand, if some process $\{X_s\}$ is obtained as a solution to some SDE of the type (17), we may be able to verify that it is a Feller process. Several sets of restrictions on μ, σ^2, γ and v are known to be sufficient for this, as found in e.g., Ch. 8 of Etheir and Kurtz (1986), Sec. 2 of Ch. IX of Revuz and Yor (1999), V.22 of Rogers and Williams

[15] For the concepts of strong and weak solutions of SDEs, see, e.g., Ch. 21 of Kallenberg (2002), or Ch. IX of Revuz and Yor (1991).

[16] Various conditions can be found in, e.g., Ch. 6 of Applebaum (2009), Ch. 5 of Either and Kurtz (1986), Ch. 21 and 23 of Kallenberg (2002), Ch. 5 of Karatzas and Shreve (1991), Ch. IV of Kunita and Watanabe (1981), Ch. IX of Revuz and Yor, (1999).

[17] For a general reference on this, see Sec. 8 in Ch. 15 of Karlin and Taylor (1981). Example 2 in Hansen and Scheinkman (1995) and Skorokhod (1961) may also be useful.

(2000), and Sec. 6.7 of Applebaum (2009). However, we do not pursue such restrictions in this paper, as they are not required for our purpose to construct a feasible testing procedure. It is our policy to start with a well-defined Feller process at hand, but not with a SDE. Regardless of this, we note that the conditions maintained in our theorems may restrict possible forms of μ, σ^2, γ, and v. For example, consider the case where $\{X_s\}$ is a diffusion process ($\gamma = 0$ and $\nu = 0$), i.e., every path of $\{X_s\}$ is almost surely continuous. In this case, since there is no isolated coffin state (by Assumption 1), the lifetime of a process may be written as $\zeta = \inf\{s \in [0,\infty) : |X_s| = \infty\}$. The condition of $\zeta = \infty$ means that the process is non-explosive. Conditions for the non-explosiveness in terms of μ and σ^2 are well-known (see, e.g., Sec. 5.5 of Karatzas and Shreve, 1991).

We subsequently present several conditions for our asymptotic results in terms of coefficients/components of the SDE (17). Before doing so, it would be worth pointing out the relationship between the components of L in (3) (the Courrège representation of $\mathcal{A}$) and those of the SDE (17). We have the following link:

$$\alpha(x) = \mu(x); \quad \beta(x) = \sigma^2(x); \quad \text{and} \quad l(x, A) = \int_{\gamma(x,z)\in A} \nu(dz). \tag{18}$$

μ and σ^2 are usually called the *drift* and *diffusion* function, respectively. However, some authors might want to use the term *drift* after a suitable adjustment. Note that the last term on the RHS of (17) is not a (local) martingale in general. If $\int_{|\gamma(x,z)|>1} \gamma(x,z)\, \nu(ds, dz) < \infty$ for each x, by letting $\bar{\mu}(x) := \mu(x) + \int_{|\gamma(x,z)|>1} \gamma(x,z)\, \nu(ds, dz)$, we can write

$$dX_s = \bar{\mu}(X_{s-})\, ds + \sigma(X_{s-})\, dW_s + \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s-}, z)\, \tilde{N}(ds, dz), \tag{19}$$

instead of (17). In this expression, the last two terms on the RHS are (local) martingales. Some authors may call this adjusted function $\bar{\mu}$ as the drift function. Note also that given the last relationship in (18) and the fact that l is a Lévy kernel, possible forms of γ and v are restricted (recall the condition in (4). Apparently, the function γ is determined only relative to the measure ν (and vice versa). From these arguments, we can see that the components in (17) (except for σ^2) may be written in various ways (one component is determined only relatively to the other ones). However, we hereafter stick to the expression (17) of $\{X_s\}$ and below provide conditions in terms of the components μ, σ, γ, and ν in (17). The form of the SDE (17) is more general and convenient (when using the Ito formula) compared to (19).[18]

[18] We consider the threshold in the last two terms on the right-hand side of (17) in terms of $\gamma(X_{s-}, z)$ with the threshold value 1. This corresponds to the form of $\mathcal{A}$ in (3), following the manner in Komatsu (1973). Some authors may prefer a different manner/expression (say, a threshold in terms of z with some other value, or a sort of smooth threshold/truncation). However, by suitable parameterization of μ, γ, and ν, we can usually check that our expression (17) may be equivalently written in some other form (see, e.g., Sec. 6.7 of Applebaum, 2009).

4.3 The asymptotic null distribution

Given the form of our test statistic and the SDE-based representation of $\{X_s\}$ in the previous subsection, we here derive the asymptotic null distribution. To develop distribution theory, we work with the following conditions:

Assumption 4 *Let $\mathcal{A}$ and $\mathfrak{D}(\mathcal{A})$ be, respectively, the infinitesimal generator and its domain of a Feller process $\{X_s\}$. $\mathcal{A}$ coincides with the following integro-differentiable operator:*

$$L\varphi(x) = \mu(x)\varphi'(x) + (1/2)\sigma^2(x)\varphi''(x) \\ + \int_{|\gamma(x,z)|>0} \left[\varphi(x+\gamma(x,z)) - \varphi(x) - \mathbf{1}_{\{|\gamma(x,z)|\le 1\}}\gamma(x,z)\varphi'(x)\right]\nu(dz), \tag{20}$$

for any test function φ in $\mathfrak{D}(\mathcal{A})$, where μ, σ^2, γ and ν satisfy the conditions in Lemma 3, and there exists some Lévy kernel l such that $l(x,A) = \int_{\gamma(x,z)\in A}\nu(dz)$ for any $A \in \mathfrak{B}(\mathbb{R}\backslash\{0\})$. Furthermore, $\mu(\cdot)$, $\sigma^2(\cdot)$ and $\gamma(\cdot,z)$ are twice continuously differentiable (for each z) with

$$|\mu(x)| + \sigma^2(x) + \int_{|\gamma(x,z)|>0} |\gamma(x,z)|^2 v(dz) \le c_1[1+\exp\{c_2|x|\}], \tag{21}$$

for some positive constants c_1 and c_2.

As discussed in Section 2, it is not generally easy to know the precise form of $\mathcal{A}$ in the whole domain $\mathfrak{D}(\mathcal{A})$. However, if some Feller process can be obtained as a solution to some SDE of the type (17), $\mathcal{A}$ must take the form of (20) for any $\varphi \in \mathfrak{D}(\mathcal{A})$, which can be checked by the Ito lemma (as in Sec. 6.7 of Applebaum, 2009). The growth condition in (21) is not restrictive at all. It is satisfied by almost all examples found in Etheir and Kurtz (1986), Sec. 2 of Ch. IX of Revuz and Yor (1999) and V.22 of Rogers and Williams (2000), Sec. 6.7 of Applebaum (2009). We also impose the following conditions on K and ρ:

Assumption 5 *The kernel function K ($\mathbb{R} \to \mathbb{R}^+$) is symmetric and twice continuously differentiable on $\mathbb{R}$ with compact support, and satisfies the following conditions: $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} zK(z)\,dz = 0$ and $\int_{\mathbb{R}} z^2K(z)\,dz < \infty$.*

Assumption 6 *The weighting function ρ ($\mathbb{R} \to \mathbb{R}^+$) is twice continuously differentiable on $\mathbb{R}$ with $\rho(x) > 0$ for any $x \in \mathbb{R}$; and there exists some constant $C_\rho > 0$ such that*

$$\sup_{x\in\mathbb{R}}\left[\rho(x) + |\rho'(x)| + |\rho''(x)| + \rho^{-1}(x)\right] < C_\rho.$$

The conditions on the kernel function K are standard except for the compactness of the support. Note that the compact-support condition is imposed for the simplicity of the proof. We may be able to work with some kernel with unbounded support, but will need to impose some tail decay condition (as in Assumption 3 of Hansen, 2008). An example of $\rho(\cdot)$ satisfying Assumption 6 is

$\rho(x) = [1+\exp\{-x\}]^{-1} + 1$, a logistic function (note that this $\rho(\cdot)$ is strictly monotone increasing (see discussions in Subsection 4.1). Given these conditions, we can now derive the asymptotic null distribution of $\hat{J}$:

Theorem 2 *Let $\{X_s\}$ be a Feller process with the infinitesimal generator $\mathcal{A}$ and its domain $\mathfrak{D}(\mathcal{A})$, satisfying the conditions in Assumptions 1-4. Suppose that the invariant density function $\pi(\cdot)$ is twice continuously differentiable on $\mathbb{R}$ and $\left|\pi^{(k)}(x)\right|$ is uniformly bounded for $k = 0, 1, 2$. Suppose also that K and ρ satisfy Assumptions 5 and 6, respectively. Let*

$$\hat{Z}(\theta) := \sqrt{T} \int_{\mathbb{R}} \hat{\Psi}(x;\theta)\, dx. \tag{22}$$

Let $n, T \to \infty$ and $\Delta, h \to 0$ with $n\Delta^2 \to 0$, $Th^4 \to 0$, and $\Delta(\log n)/h \to 0$. Then, if the null hypothesis H_0 holds,
(i) there exists a mean-zero Gaussian process $\{Z_0(\theta)\}_{\theta\in\Theta}$ whose covariance kernel is $\Lambda_0(\theta_1, \theta_2)$, such that $\left\{\hat{Z}(\theta)\right\}_{\theta\in\Theta}$ converges weakly to $\{Z_0(\theta)\}_{\theta\in\Theta}$ in $\mathbf{C}(\Theta)$ (the space of continuous functions on Θ);
(ii) $\hat{\lambda}(\theta) \xrightarrow{P} \lambda_0(\theta) := \Lambda_0(\theta, \theta)$ uniformly over $\theta \in \Theta$.

The proof is provided in the Appendix. Note that the convergence rate of $\hat{Z}(\theta)$ under the null is independent of that of the smoothing parameter h. This is due to the fact the convergence rate of the integral of the kernel-based estimator is faster than that of the original kernel-based estimator (a similar phenomenon can be found in Vanhems, 2006). Given the result of the theorem, we have

$$\hat{J} \Longrightarrow J_0 := \int_{\Theta} Z_0^2(\theta)\, d\theta \Big/ \int_{\Theta} \lambda_0(\theta)\, d\theta,$$

by the continuous mapping theorem. While this limit of J_0 is case dependent (non-pivotal), we can find an upper bound of J_0 which is independent of any unknown objects. By Mercer's theorem with the aid of a certain linear programming problem (see Sec. 6 of Bierens and Ploberger, 1997), we have

$$\lim \Pr\left[\hat{J} > c\right] \leq \Pr\left[\bar{W} > c\right],$$

where $\bar{W} := \sup_{m\geq 1} m^{-1} \sum_{j=1}^{m} \varepsilon_j^2$ and $\{\varepsilon_j\}_{j\geq 1}$ is a sequence of i.i.d. random variables with $\varepsilon_j \sim N(0,1)$. Since we can obtain quantiles of $\bar{W}$ (through a Monte Carlo simulation), we can implement a conservative test. For example, Bierens and Ploberger (1997) report

$$\Pr\left[\bar{W} > 3.23\right] = 0.10; \quad \Pr\left[\bar{W} > 4.26\right] = 0.05; \quad \text{and} \quad \Pr\left[\bar{W} > 6.81\right] = 0.01.$$

The use of these conservative bounds obviously introduces some size distortions of the test. We investigate the effects due to these upper bounds by Monte Carlo experiments in the next section.

Note that given the identification and weak-convergence results (in Theorems 1 and 2, respectively), we might be able to use $\hat{N} := \int_{\Theta} |\hat{Z}(\theta)|^2 d\theta$ for our test statistic, while the covariance kernel of $\{Z_0(\theta)\}$

depends on several unknown objects and the limit object $\int_\Theta |Z_0(\theta)|^2 d\theta$ is non-pivotal, which makes it impossible to tabulate critical values. Even when tabulation is impossible, we might be able to estimate/approximate critical values. For example, we can construct a nonparametric estimator of the covariance kernel $\Lambda_0(\theta_1, \theta_2)$, and verify its consistency under the null hypothesis.[19] Then, by using the estimated covariance kernel, we could simulate the null distribution of $\int_\Theta |Z_0(\theta)|^2 d\theta$ and then conduct a (asymptotically) size-correct test. However, this approach may lead to the loss of power/consistency of the test. This is because the convergence rate of $\hat{Z}(\theta)$ is *different* under the null and alternative hypotheses, and we cannot necessarily expect that $\hat{Z}(\theta) \to \infty$ (and $\hat{N} \to \infty$) under the alternative. If $\{X_s\}$ is the Brownian motion, for example, we would only have $\hat{Z}(\theta) = O_P(1)$ and $\hat{N} = O_P(1)$ (see discussions on the convergence rates in the generalized LLN in the next subsection).[20] The problem here is that $\sqrt{T}$ as in (16) and (22) is not an appropriate normalization rate under the alternative (such a rate is generally unknown, unfortunately).

By the same reasoning, it is also uncertain if we could get through the problem by using the so-called conditional Monte Carlo (or p-value) approach (see Hansen, 1996; De Jong, 1996), or Escanciano and Jacho-Chávez's (2010) approach to estimate eigenelements of the covariance kernel. Additionally, for the validation of these approaches, it seems necessary to impose a certain mixing (or weak-dependence) condition. In the light of our testing purpose, it should be reasonable to maintain the same conditions under both the null and alternative hypotheses (other than these hypotheses themselves). If we imposed some mixing condition under both the hypotheses, we would have the class of alternative processes essentially empty. Although the mixing is in principle a different concept from the stationarity, they are quite interrelated.

As another approach, one might think of using some sort of bootstrap. However, recall that our null restriction is fully nonparametric. Therefore, it is not obvious how to construct a bootstrap analog of $\hat{J}$ (or $\hat{N}$) which incorporates such nonparametric restriction. If we use a certain sort of nonparametric bootstrap without the null restriction, we conjecture that a bootstrap analog of $\hat{J}$ tends to ∞ under the alternative hypothesis, which results in no power/consistency of the test.

[19] We might be able to estimate $\Lambda_0(\theta_1, \theta_2)$ by

$$\hat{\Lambda}(\theta_1, \theta_2) := T^{-1} \sum_{i=0}^{n-1} \left| \int_{\mathbb{R}} K_h(X_{i\Delta} - x) \rho^{-1}(x) dx \rho(X_{i\Delta}) \right|^2$$
$$\times \left[\phi\left(X_{(i+1)\Delta}; \theta_1\right) - \phi(X_{i\Delta}; \theta_1)\right] \left[\phi\left(X_{(i+1)\Delta}; \theta_2\right) - \phi(X_{i\Delta}; \theta_2)\right].$$

[20] Under the alternative, it will generally hold that $\hat{\Lambda}(\theta_1, \theta_2) = o_P(1)$ ($\hat{\Lambda}(\theta_1, \theta_2)$ is defined in the previous footnote), and therefore, simulated critical values (based on the estimate of $\Lambda_0(\theta_1, \theta_2)$) may also be expected to approach zero as $T \to \infty$. If this is the case, we conjecture that the test using $\hat{N}$ and simulated critical values may have some power/consistency property. However, we leave the verification of this conjecture to future research, as it will require some extra work.

4.4 The asymptotic power property

We here show that our testing procedure with the test statistic $\hat{J}$ is consistent and has non-trivial power against any (fixed) alternative null recurrent process.

Theorem 3 *Suppose the same conditions as in Theorem 2. Then, if the alternative hypothesis H_1 holds, there exist some constants $\beta \in (0,1)$ and $C > 0$ such that*

$$\hat{J}/T^{\beta} \geq C \to \infty,$$

with probability approaching to 1 *(as $T \to \infty$).*

The proof of this theorem uses a generalized LLN for nonstationary (null recurrent) processes. The divergnce rate of the test statistic $\hat{J}$ is determined by β. This factor β corresponds to the divergence/convergence rate in the generalized LLN, i.e., β satisfies

$$\int_0^T g(X_s)\, ds = O_P\left(T^{\beta+\varepsilon}\right), \tag{23}$$

for a bounded function g with $\int g(x)\,\pi(x)\,dx < \infty$ (for any arbitrarily small $\varepsilon > 0$). For example, if $\{X_s\}$ is a Brownian motion ($X_s = W_s$), then we have $\int_0^T g(X_s)\,ds = O_P\left(T^{1/2}\right)$ and $\beta = 1/2$. In the Markov chain terminology, a Markov process satisfying a discrete-time counterpart of (23) is said to be β-recurrent (see, e.g., Karlsen and Tjøstheim, 2001). For our continuous-time Markov case, the existence of β satisfying (23) is guaranteed for any null recurrent process (Sec. 3.3 of Höpfner and Löcherbach, 2003).

5 Monte Carlo Results

In this section, we examine finite-sample size and power properties of the proposed test. First, see the size performance (and the conservativeness of the upper bound approximation), we consider a simulation study with the following data-generating processes:

Model 1: The Ornstein-Uhlenbeck (OU) process, whose stationarity is drift-induced,

$$dX_s = \lambda(m - X_s)\,ds + \sigma dW_s,$$

with $(\lambda, m, \sigma^2) = (0.85837, 0.089102, 0.0021854)$, taken from Aït-Sahalia's (1996, Table III in p. 542) estimates for the seven-day Eurodollar rate data.

Model 2: Aït-Sahalia's (1999) nonlinear process with drift-induced stationarity:

$$dX_s = \left(\alpha_{-1}X_s^{-1} + \alpha_0 + \alpha_1 X_s + \alpha_2 X_s^2\right) ds + \sigma X_s^{3/2} dW_s,$$

with $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma) = (0.000693, -0.0347, 0.676, -4.059, 0.84214)$, taken from Aït-Sahalia's (1999, Table VI in p. 1389) estimates for the monthly Federal Funds rate data.

Model 3: Bibby and Sørensen's (1997) hyperbolic diffusion process:

$$dX_s = \sigma \exp\left\{\frac{1}{2}\left(\alpha\sqrt{\delta^2 + (X_s - \mu)^2} - \beta(X_s - \mu)\right)\right\} dW_s,$$

where $(\alpha, \beta, \delta, \mu, \sigma) = (4.4875, -3.8412, 1.1949, 7.2915, 0.0047)$, taken from Bibby and Sørensen's (1997, Table 1 in p. 35) estimates for Baltica stock price data.

We measure time in years and consider the following two observation intervals: $\Delta = 1/12$ and $1/252$, which correspond to sampling every month and every day, respectively.[21] For each Δ, we consider two cases: $T = 20$ and 40. In order to simulate data, we use the exact simulation scheme with the random number generator of the normal distribution for the OU process (see, e.g., p. 456 of Pritsker, 1998). For the Aït-Sahalia nonlinear process, we employ the Euler-Maruyama discretization scheme (see Higham, Mao and Stuart, 2003) with the discretization step $d = \underline{\Delta}/100$, where $\underline{\Delta}^{-1} = 252$ corresponds to the highest sampling frequency used in this simulation study. For the Bibby and Sorensen hyperbolic diffusion process, we use the strong Taylor scheme of order 1.5 with the discretization step $d = \underline{\Delta}/25$ (see Sec. 3 of Bibby and Sørensen, 1997).[22]

Throughout this experiment, we let $\Xi = [-1, 1]$ and

$$\rho^{-1}(x) = 1 + G(x) = 1 + [1 + \exp\{-x\}]^{-1},$$

where $G(x)$ is the cumulative distribution function of a logistic random variable. For this choice $\rho^{-1}(x)$, we can check that $\rho(x)$ is strictly increasing with satisfying the conditions in Assumption 6. We compute the integrations with respect to x and θ by Monte-Carlo integrations based on a so-called low-discrepancy sequence (the Halton sequence), where we outline our integration method in the Appendix. We use the Epanechnikov kernel with the bandwidth parameter h chosen according to $h = 1.06\hat{s}n^{-1/5}$ (the so-called rule of thumb in the density estimation i.i.d. data), where $\hat{s}$ is the standard deviation of the observations.[23] This choice of h satisfies the conditions in Theorems 2 and 3.

By using the upper bounds of the 5% and 10% critical values, we compute the percentage of rejections of the null hypothesis H_0 based on 400 replications, reported in Table 1. From the results in Table 1, we can see that the test has some size distortions. In particular, for $\Delta = 1/252$ and $T = 40$, the test tends to exhibit more rejection rates than the nominal sizes. This is an expected phenomenon since critical

[21] Roughly, 252 corresponds to the numbers of business days in a year.

[22] To simulate data with the discretization schemes, we start from the initial value $X_0 = 0.0717$ (the mean of the monthly Federal Funds rates) for the Aït-Sahalia nonlinear process, and $X_0 = \mu + \delta\beta/\sqrt{\alpha^2 - \beta^2}$ (the mode setting of the invariant distribution of X_s) for the Bibby and Sørensen hyperbolic diffusion process. We simulate a trajectory of $T \times 1.2$ years, and discard the first $T \times 0.2$-year fraction of each trajectory (to make the effect of the initial value negligible).

[23] To check the sensitivity of the proposed test with respect to the choice of bandwidth, we also considered some other bandwidths, say, $h = c\hat{s}n^{-1/5}$ with setting different values of $c : c = 1/4, 1/2, 2,$ and 4. However, we obtained similar size and power properties for all bandwidths we used. This may be explained by the fact that the convergence rate of the test statistic is independent of h.

values used are conservative ones. For the other cases, we observe over-rejection tendencies, which may be due to an artifact of small samples.

We also simulate the following models to examine the power property of the test in finite samples:

Model 4: The standard Brownian motion $X_s = W_s$.

Model 5: Höpfner and Kutoyants's (2003) model:

$$dX_t = -\frac{\theta X_s}{1 + X_s^2} ds + \sigma dW_s,$$

where we set $(\theta, \sigma) = (1/4, 1)$.

The Höpfner and Kutoyants model is simulated by using the same method as for Model 2 (we set the initial value $X_0 = 0$). By using the same settings as above, we also computed the percentage of rejections of the null hypothesis H_0 based on 400 replications (Table 2). The results reported in Table 2 suggest that our test has some non-trivial power when Δ is small and T is large.

Table 1: Percentage of rejections of the **true** H_0

	Model 1		Model 2		Model 3	
Nominal Size	5%	10%	5%	10%	5%	10%
$\Delta = 1/12$, $T = 20$ ($n = 240$)	0.0750	0.1425	0.0825	0.1750	0.1325	0.1800
$\Delta = 1/12$, $T = 40$ ($n = 480$)	0.0600	0.1675	0.0950	0.1900	0.0875	0.2125
$\Delta = 1/252$, $T = 20$ ($n = 5040$)	0.0575	0.1300	0.0825	0.1575	0.0900	0.1750
$\Delta = 1/252$, $T = 40$ ($n = 10080$)	0.0250	0.0825	0.0425	0.0750	0.0550	0.0975

Table 2: Percentage of rejections of the **false** H_0

	Model 4		Model 5	
Nominal Size	5%	10%	5%	10%
$\Delta = 1/12$, $T = 20$ ($n = 240$)	0.1125	0.1925	0.0825	0.1750
$\Delta = 1/12$, $T = 40$ ($n = 480$)	0.1950	0.3275	0.1700	0.2725
$\Delta = 1/252$, $T = 20$ ($n = 5040$)	0.2450	0.3600	0.2625	0.3100
$\Delta = 1/252$, $T = 40$ ($n = 10080$)	0.3750	0.4825	0.3050	0.4300

6 Conclusion

We have proposed a new statistical testing procedure to examine the stationarity property of continuous-time Markov processes based on the restriction through the infinitesimal generator. Our test is based on two novel propositions: (i) a new theorem to identify the stationarity property using the nuisance parameter approach; (ii) asymptotic theory for the proposed test statistic. The identification scheme is fully nonparametric and does not rely on the concept of the unit root or integration. It allows us to assess the generic stationarity property of time series processes, and can serve as a new alternative to DF and KPSS type tests. The asymptotic theory contained in this paper is based on the Markov regeneration technique and is derived without imposing any exploit mixing condition.

A Appendix

A.1 Proofs

Proof of Lemma 2. Consider a smooth truncation function (indexed by $\varepsilon > 0$) as follows:

$$\tau_\varepsilon(x) := \begin{cases} 1 & \text{if } |x| \leq \varepsilon; \\ \exp\{-\exp\{-1/(|x|-\varepsilon)^2\}/(|x|-\varepsilon-1)^2\} & \text{if } |x| \in (\varepsilon, \varepsilon+1); \\ 0 & \text{if } |x| \geq \varepsilon + 1. \end{cases} \tag{24}$$

This function is infinitely differentiable and compactly supported.[24] Fix any arbitrary $\varphi \in \hat{\mathbf{C}}^k(\mathbb{R})$ and let $\eta_\varepsilon(x) := \varphi(x)\tau_\varepsilon(x)$ for each $\varepsilon > 0$. When φ is k-times continuously differentiable (with some $k \geq 0$), η_ε is also so. In this case, noting the functional form of $\tau_\varepsilon(x)$, as well as the fact that $\varphi^{(i)} \in \hat{\mathbf{C}}(\mathbb{R})$ for any $i \leq k$ (this holds since $\varphi \in \hat{\mathbf{C}}(\mathbb{R})$), we have

$$\sum\nolimits_{i=0}^{k} \sup_{x\in\mathbb{R}} \left|\varphi^{(i)}(x) - \eta_\varepsilon^{(i)}(x)\right| \to 0 \ \text{ as } \ \varepsilon \to \infty. \tag{25}$$

By the truncation, η_ε is compactly supported. On the other hand, by Lemma 4, whose statement and proof are provided below, for each $\varepsilon > 0$, there exists a sequence of functions $\left\{H_{\tilde{J}}^\varepsilon(x)\right\}$ such that $H_{\tilde{J}}^\varepsilon(x) = L_{\tilde{J}}^\varepsilon(x)\mathfrak{w}(x)$, $L_{\tilde{J}}^\varepsilon(x)$ is a polynomial function of degree at most $\tilde{J}-1$, and

$$\sum\nolimits_{i=0}^{k} \sup_{x\in\mathbb{R}} \left|\eta_\varepsilon^{(i)}(x) - H_{\tilde{J}}^\varepsilon(x)\right| \to 0 \ \text{ as } \ \tilde{J} \to \infty, \tag{26}$$

where we note that $\tilde{J}$ depends on ε, i.e, $\tilde{J} = \tilde{J}(\varepsilon)$. By (25) and (26), we can construct a sequence of functions $\{H_J(\cdot)\}$ satisfying the conditions in the lemma, completing the proof. ■

Lemma 4 *Let η be an arbitrary function in $\hat{\mathbf{C}}_K(\mathbb{R})$ $(k \geq 0)$ which is k-times continuously differentiable functions. Then, for each η, there exists a sequence of functions, $\{L_J(\cdot) : J = k+1, k+2, \dots\}$, such that each $L_J(\cdot)$ is a polynomial function of degree at most $J-1$, and*

$$\sum\nolimits_{i=0}^{k} \sup_{x\in\mathbb{R}} \left|\eta^{(i)}(x) - H_J^{(i)}(x)\right| \to 0 \ \text{ as } \ J \to \infty,$$

where $H_J(x) := L_J(x)\mathfrak{w}(x)$, and $f^{(i)}$ and $H_J^{(i)}$ are the i-th order derivatives of φ and H_J, respectively.

Proof of Lemma 4. Let

$$\eta(x) = [\eta(x)/\mathfrak{w}(x)] \times \mathfrak{w}(x) =: f(x) \times \mathfrak{w}(x),$$

where $f(x)$ is well-defined over $x \in \mathbb{R}$ since $\mathfrak{w}(x) > 0$ for any $x \in \mathbb{R}$. For any continuous function g, define the following set of functions:

$$\mathfrak{C}^k(\mathbb{R}; g) := \{q\,(: \mathbb{R} \to \mathbb{R}) \mid q \text{ is } k\text{-times continuously differentiable on } \mathbb{R};$$
$$q^{(i)} g \in \hat{\mathbf{C}}(\mathbb{R}) \ \text{ for } i = 0, \dots, k\},$$

[24]The form in (24) is only one example, and we can think of some other functional form satisfying the smoothness and compact-support conditions.

where $\hat{\mathbf{C}}(\mathbb{R})$ is the set of continuous functions on $\mathbb{R}$ which vanish at infinity (as defined in Section 2). Now, let $w_c(x) := \exp\{-cx^2\}$ for an arbitrary constant $c > 0$, and consider the set of functions $\mathfrak{C}^k(\mathbb{R}; w_c)$, where we note that $\mathfrak{C}^k(\mathbb{R}; w_{1/2}) = \mathfrak{C}^k(\mathbb{R}; \mathfrak{w})$. Since $\eta \in \mathbf{C}_K^k(\mathbb{R})$ and the support of f is compact, it holds that $\left|f^{(i)}(x)\right| \to 0$ as $|x| \to 0$ for any $i \le k$. Therefore, $f \in \mathfrak{C}^k(\mathbb{R}; w_c)$ for any $c > 0$.

Now, let Π_J denote the set of polynomial functions of degree at most J. For a function g, we also define the following quantity:

$$\mathbf{E}_J(g)_{w_c} := \inf_{p\in\Pi_J} \sup_{x\in\mathbb{R}} |w_c(x)[g(x) - p(x)]|.$$

Fix any (arbitrary) $c > 0$. Then, by inequalities (2) and Corollary 1 of Balázs (2004), we can construct a sequence of polynomial functions $\{L_J(\cdot) : J = k+1, k+2, \dots\}$ (based on the Lagrange interpolation method) such that each L_J is a polynomial function of degree at most $J-1$, and for some constants $c_3, c_4 > 0$,

$$\sum_{i=0}^{k} \sup_{x\in\mathbb{R}} \left|f^{(i)}(x) - L_J^{(i)}(x)\right| w_c(x) \le \sum_{i=0}^{k} \alpha_{i,k} \left(\frac{c_3 [J+1-k]^{1/(1+c_4)}}{J+1-k}\right)^{k-i} \mathbf{E}_{J-k+1}\left(f^{(k)}\right)_{w_c} \log J, \tag{27}$$

where each $\alpha_{i,k}$ is a constant depending only on i, k and w_c. Since $f^{(k)} \in \mathfrak{C}^k(\mathbb{R}; w_c)$, it holds that

$$\mathbf{E}_{J-k+1}\left(f^{(k)}\right)_{w_c} \to 0 \text{ as } J \to \infty, \tag{28}$$

which follows from arguments in p. 100 of Szabados (1997). By (27) and (28), there exists a sequence of functions $\{L_J(\cdot) : J = k+1, k+2, \dots\}$ such that each L_J is a polynomial function of degree at most $J-1$ and

$$\sum_{i=0}^{k-1} \sup_{x\in\mathbb{R}} \left|f^{(i)}(x) - L_J^{(i)}(x)\right| w_c(x) \to 0 \text{ as } J \to \infty. \tag{29}$$

Given (29), we now prove the statement of the lemma. If $k = 0$, (10) holds obviously by (29) with $c = 1/2$. If $k = 1$, let any $c \in (0, 1/2)$ and consider a sequence $\{L_J(\cdot)\}$ satisfying (29). In this case,

$$\begin{aligned}\left|\varphi^{(1)}(x) - H_J^{(1)}(x)\right| &= \left|f^{(1)}(x)\mathfrak{w}(x) - f(x)\mathfrak{w}^{(1)}(x) - L_J^{(1)}(x)\mathfrak{w}(x) - L_J(x)\mathfrak{w}^{(1)}(x)\right| \\ &\le \left|f^{(1)}(x) - L_J^{(1)}(x)\right|\mathfrak{w}(x) + |f(x) - L_J(x)|\left|\mathfrak{w}^{(1)}(x)\right| \\ &\le C\left|f^{(1)}(x) - L_J^{(1)}(x)\right| w_c(x) + C|f(x) - L_J(x)| w_c(x),\end{aligned} \tag{30}$$

where the last inequality follows from the fact that for each positive integer k, for any $\tilde{c} \in (0, c)$, there exists some positive constant C such that

$$\max_{i\in\{0,1,\dots,k\}} \left|\mathfrak{w}^{(i)}(x)\right| \le C w_c(x). \tag{31}$$

By (29), the RHS of (30) tends to zero uniformly as $J \to \infty$, i.e.,

$$\sup_{x\in\mathbb{R}} \left|\varphi^{(1)}(x) - H_J^{(1)}(x)\right| \to 0 \text{ as } J \to \infty,$$

which, together with the result for $k = 0$, gives the desired result. For the case where $k \geq 2$, the proof can be done analogously by using the product differentiation formula and (31), and we omit details. The proof is completed. $\blacksquare$

Proof of Theorem 1. The "if" part is obvious. We prove the "only if" part. Now, suppose that $\int_{\mathbb{R}} \mathcal{A}\varphi(x)\,\pi(x)\,dx \neq 0$ for some $\varphi \in \mathfrak{D}(\mathcal{A})$. Now, by Lemma 2, we can construct some sequence of weighted polynomial functions $\{H_J(x)\}$ such that $H_J(x) = L_J(x)\,\mathfrak{w}(x)$, $L_J(x)$ is a polynomial function of degree at most $J-1$, and

$$\sum\nolimits_{i=0}^{k} \sup\nolimits_{x\in\mathbb{R}} \left|\varphi^{(i)}(x) - H_J(x)\right| \to 0 \quad \text{as} \quad J \to \infty. \tag{32}$$

By the form of H_J, we can write

$$H_J(x) = \sum\nolimits_{l=0}^{J-1} \gamma_l g_l(x),$$

where $g_l(x) = x^l \mathfrak{w}(x)$ and $\{\gamma_l\}$ is a sequence of some constant coefficients. By the linearity of $\mathcal{A}$ and the condition (i) of Assumption 3, it holds that $H_J(\cdot) \in \mathfrak{D}(\mathcal{A})$ for any J, and therefore,

$$\int_{\mathbb{R}} \mathcal{A}H_J(x)\,\pi(x)\,dx = \sum\nolimits_{l=0}^{J} \gamma_l \int_{\mathbb{R}} \mathcal{A}g_l(x)\,\pi(x)\,dx, \tag{33}$$

If $\left|\int_{\mathbb{R}} \mathcal{A}\varphi(x)\,\pi(x)\,dx\right| < \infty$, (ii) of Assumption 3 and the result (32) imply that for J large enough, $\int_{\mathbb{R}} \mathcal{A}H_J(x)\,\pi(x)\,dx \neq 0$. If $\left|\int_{\mathbb{R}} \mathcal{A}\varphi(x)\,\pi(x)\,dx\right| = \infty$, consider the set $E_N := [-N, N]$ for a positive integer N. In this case, by the continuity of $\mathcal{A}\varphi(x)\,\pi(x)$, it holds that $\left|\int_{E_N} \mathcal{A}\varphi(x)\,\pi(x)\,dx\right| < \infty$ for any N, but for N large enough, we can obtain $\int_{E_N} \mathcal{A}\varphi(x)\,\pi(x)\,dx \neq 0$. And, by (ii) of Assumption 3 and the result (32), we also have $\int_{E_N} \mathcal{A}H_J(x)\,\pi(x)\,dx \neq 0$. On the other hand, since $\left|\int_{\mathbb{R}} \mathcal{A}g_l(x)\,\pi(x)\,dx\right| < \infty$ holds for any l, it also holds that $\left|\int_{\mathbb{R}} \mathcal{A}H_J(x)\,\pi(x)\,dx\right| < \infty$. By letting N be large enough, we can let $\int_{\mathbb{R}\setminus E_N} \mathcal{A}H_J(x)\,\pi(x)\,dx$ be arbitrary small. Therefore, we also have $\int_{\mathbb{R}} \mathcal{A}H_J(x)\,\pi(x)\,dx \neq 0$.

Given $\int_{\mathbb{R}} \mathcal{A}H_J(x)\,\pi(x)\,dx \neq 0$, (33) implies that there exists some $l^*\ (\leq J)$ such that

$$\int_{\mathbb{R}} \mathcal{A}g_{l^*}(x)\,\pi(x)\,dx \neq 0. \tag{34}$$

Now, observe that

$$\phi(x;\theta) = \exp(\theta x)\,\mathfrak{w}(x) = \sum\nolimits_{k=0}^{\infty} \left(\theta^k/k!\right) x^k \mathfrak{w}(x) = \lim_{J\to\infty} \phi_J(x;\theta),$$

where

$$\phi_J(x;\theta) := \sum\nolimits_{l=0}^{J} \left(\theta^l/l!\right) x^l \mathfrak{w}(x) = \sum\nolimits_{l=0}^{J} \left(\theta^l/l!\right) g_l(x).$$

By arguments similar to those in the proof of Lemma 4, the simultaneous uniform convergence of $\phi_J(\cdot;\theta)$ and its derivatives up to the $\bar{k}$-th order occurs ($\bar{k}$ may be arbitrary large). That is, for each θ,

$$\sum\nolimits_{i=0}^{\bar{k}} \sup\nolimits_{x\in\mathbb{R}} \left|\phi^{(i)}(x;\theta) - \phi_J^{(i)}(x;\theta)\right| \to 0 \quad \text{as} \quad J \to \infty, \tag{35}$$

Then, by the linearity of the integral and $\mathcal{A}$,

$$\int_{\mathbb{R}} \mathcal{A}\phi_J(x;\theta)\,\pi(x)\,dx = \sum\nolimits_{l=0}^{J} \left(\theta^l/l!\right) \int_{\mathbb{R}} \mathcal{A}g_l(x)\,\pi(x)\,dx.$$

Now, let $k = \bar{k}$ in (35). In this case, by (ii) of Assumption 3, the limit of $\int_{\mathbb{R}} \mathcal{A}\phi_J(x;\theta)\,\pi(x)\,dx$ is well-defined and

$$\int_{\mathbb{R}} \mathcal{A}\phi(x;\theta)\,\pi(x)\,dx = \sum_{l=0}^{\infty} \left(\theta^l/l!\right) \int_{\mathbb{R}} \mathcal{A}g_l(x)\,\pi(x)\,dx, \tag{36}$$

for each θ (note that $\mathcal{A}\phi(\cdot;\theta)$ is well-defined since $\phi(\cdot;\theta) \in \mathfrak{D}(\mathcal{A})$).

By (36), we have checked that the term-wise operation of the integral and $\mathcal{A}$ to $\phi(\cdot;\theta)$ is permitted. Let

$$L(\theta) = \sum_{l=0}^{\infty} \left(\theta^l/l!\right) \int_{\mathbb{R}} \mathcal{A}g_l(x)\,\pi(x)\,dx,$$

where $L(\theta)$ is a power series of θ whose radius of convergence is ∞. Therefore, the term-wise differentiation of $L(\theta)$ (at any θ) is also permitted:

$$\frac{d^{l^*}}{d\theta^{l^*}} L(\theta) = \sum_{l=0}^{\infty} \frac{d^{l^*}}{d\theta^{l^*}} \left(\theta^l/l!\right) \int_{\mathbb{R}} \mathcal{A}g_l(x)\,\pi(x)\,dx.$$

Letting $\theta \to 0$, the RHS converges to $\int_{\mathbb{R}} \mathcal{A}g_{l^*}(x)\,\pi(x)\,dx$. This and (34) imply that for some θ (in the neighborhood of zero), $L(\theta) \neq 0$. Noting the continuity of $L(\theta)$, we obtain the desired result. The proof is completed. ■

Proof of Lemma 3. For any Feller process $\{X_s\}$, we can write

$$\varphi(X_t) - \varphi(X_0) = \int_0^t \mathcal{A}\varphi(X_s)\,ds + \mathcal{M}_t^{\varphi}, \tag{37}$$

for any test function $\varphi \in \mathfrak{D}(\mathcal{A})$, where $\{\mathcal{M}_t\}$ is a martingale for each $\theta \in \Theta$. The validity of this expression can be shown by Lemma 19.21 of Kallenberg (2002). Suppose that $\mathfrak{D}(\mathcal{A})$ contains $\mathbf{C}_K^{\infty}(\mathbb{R})$. In this case, recall that $\mathcal{A} = L$ on the space of $\mathbf{C}_K^{\infty}(\mathbb{R})$, as in (3). Let $\mathcal{N}_t^{\varphi} := \varphi(X_t) - \varphi(X_0) - \int_0^t L\varphi(X_s)\,ds$. Then, $\{\mathcal{N}_t^{\varphi}\}$ is a martingale for any $\varphi \in \mathbf{C}_K^{\infty}(\mathbb{R})$. This means that $\{\mathcal{N}_t^{\varphi}\}$ is also a martingale for any $\varphi \in \mathbf{C}_b^2(\mathbb{R})$ by Theorem 1.1 of Stroock (1975), where $\mathbf{C}_b^2(\mathbb{R})$ is the space of bounded and twice continuously differentiable functions on $\mathbb{R}$ whose derivatives are also bounded.[25] Now, by Theorem 2.2 of Komatsu (1973), the conclusion follows. ■

Proof of Theorem 2. Since $\phi(x;\theta)$ is uniformly bounded (over x and θ), ν satisfies the restriction in (18) and l is the Levy measure, it holds that for each x,

$$\int_{|\gamma(X_{s-},z)|>1} \left[\phi(x + \gamma(x,z);\theta) - \phi(x;\theta)\right] \nu(dz) < \infty.$$

Then, by the Ito formula, we can write

$$\begin{aligned} \phi\left(X_{(i+1)\Delta};\theta\right) - \phi(X_{i\Delta};\theta) &= \int_{i\Delta}^{(i+1)\Delta} g_\theta(X_{s-})\,ds + \int_{i\Delta}^{(i+1)\Delta} \phi'(X_{s-};\theta)\,\sigma(X_{s-})\,dW_s \\ &\quad + \int_{i\Delta}^{(i+1)\Delta} \int_{\mathbb{R}\setminus\{0\}} \left[\phi(X_{s-} + \gamma(X_{s-},z);\theta) - \phi(X_{s-};\theta)\right] \tilde{N}(ds,dz), \end{aligned} \tag{38}$$

[25] This implies that $\{X_s\}$ is a solution to the martingale problem associated to L on $\mathbf{C}_b^2(\mathbb{R})$.

where

$$g_\theta(x) := L\phi(x;\theta) = \left[\mu(x)\phi'(x;\theta) + (1/2)\sigma^2(x)\phi''(x;\theta)\right] + \int_{\mathbb{R}\backslash\{0\}} \left[\phi(x+\gamma(x,z);\theta) - \phi(x) - \gamma(x,z)\phi'(x;\theta)\right]\nu(dz),$$

where L is the integro-differential operator defined in (3). We note that the last two terms on the RHS of (38) are martingales whose moments of any order exist (recall Assumption 4 and the form of $\phi(x;\theta)$).

Given the expression (38), we can write

$$\int_{\mathbb{R}} \hat{\Psi}(x;\theta)\,dx = Q_n(\theta) + R_n(\theta),$$

where

$$\begin{aligned}
Q_n(\theta) &:= T^{-1/2}\sum_{i=0}^{n-1}\int_{\mathbb{R}} K_h(X_{i\Delta}-x)\left[\rho(X_{i\Delta})/\rho(x)\right]dx \int_{i\Delta}^{(i+1)\Delta} g_\theta(X_{s-})\,ds;\\
R_n(\theta) &:= T^{-1/2}\int_0^T \int_{\mathbb{R}} K_h(X_{i\Delta}-x)\left[\rho(X_{i\Delta})/\rho(x)\right]dx\phi'(X_{s-};\theta)\,\sigma(X_{s-})\,dW_s\\
&+ T^{-1/2}\int_0^T \int_{\mathbb{R}} K_h(X_{i\Delta}-x)\left[\rho(X_{i\Delta})/\rho(x)\right]dx\\
&\times \int_{\mathbb{R}\backslash\{0\}} \left[\phi(X_{s-}+\gamma(X_{s-},z);\theta) - \phi(X_{s-};\theta)\right]\tilde{N}(ds,dz).
\end{aligned}$$

We first show that $Q_n(\theta) = o_P(1)$ uniformly over θ (and therefore the asymptotic distribution is determined by $R_n(\theta)$. For this purpose, we also consider the following decomposition:

$$Q_n(\theta) = Q_{n,1}(\theta) + Q_{n,2}(\theta),$$

where two components on the RHS are defined as follows (note that $\int g_\theta(x)\pi(x)\,dx = 0$ under the null hypothesis):

$$\begin{aligned}
Q_{n,1}(\theta) &:= T^{-1/2}h^{-1}\sum_{i=0}^{n-1}\int_{\mathbb{R}} K((X_{i\Delta}-x)/h)\left[\rho(X_{i\Delta})/\rho(x)\right]dx\int_{i\Delta}^{(i+1)\Delta}\left[g_\theta(X_{s-}) - g_\theta(X_{i\Delta})\right]ds;\\
Q_{n,2}(\theta) &:= T^{-1/2}(1/nh)\sum_{i=0}^{n-1}\int_{\mathbb{R}} K((X_{i\Delta}-x)/h)\left[\rho(X_{i\Delta})/\rho(x)\right]\left[g_\theta(X_{i\Delta}) - g_\theta(x)\right]dxds;\\
Q_{n,3}(\theta) &:= T^{1/2}(1/nh)\sum_{i=0}^{n-1}\int_{\mathbb{R}} K((X_{i\Delta}-x)/h)\left[\rho(X_{i\Delta})/\rho(x)\right]g_\theta(x)\,dx - \int g_\theta(x)\pi(x)\,dx.
\end{aligned}$$

By using the Ito formiula, as well as the uniform boundedness and continuity of g_θ and ρ (and their derivatives), we can show that

$$Q_{n,1}(\theta) = O_p(\sqrt{T\Delta}) \text{ uniformly over } \theta.$$

As for $Q_{n,2}(\theta)$, we look at

$$Q_{n,2}(\theta) = \sqrt{T}(1/n)\sum_{i=0}^{n-1}\int_{\mathbb{R}} K(q)\left[\rho(X_{i\Delta})/\rho(X_{i\Delta}-qh)\right]\left[g_\theta(X_{i\Delta}) - g_\theta(X_{i\Delta}-qh)\right]dq,$$

where the equality holds by changing variables with $q = (X_{i\Delta} - x)/h$. Then, by the standard arguments Taylor approximation for kernel-based estimators, we have

$$Q_{n,2}(\theta) = O(\sqrt{T}h^2) \text{ uniformly over } \theta.$$

To find the order of $Q_{n,3}(\theta)$, we also look at

$$\begin{aligned} Q_{n,3}(\theta) &= T^{1/2}(1/nh)\sum_{i=0}^{n-1}\int [Y_{i\Delta}(x) - E[Y_{i\Delta}(x)]]\, g_\theta(x)\, dx \\ &+ T^{1/2}(1/nh)\sum_{i=0}^{n-1}\int_{\mathbb{R}} [E[K((X_{i\Delta} - x)/h)[\rho(X_{i\Delta})/\rho(x)]] - \pi(x)]\, g_\theta(x)\, dx, \end{aligned} \tag{39}$$

where $Y_{i\Delta}(x) := K((X_{i\Delta} - x)/h)[\rho(X_{i\Delta})/\rho(x)]$. By the same arguments as before, we can show that the second term on the RHS (39) is $O(\sqrt{T}h^2)$ (uniformly over θ). To find the the order of the first term, we use the following result:

$$(1/nh)\sum_{i=0}^{n-1}\int [Y_{i\Delta}(x) - E[Y_{i\Delta}(x)]]\, g_\theta(x)\, dx = O_P(\sqrt{(\log n)/nh}),$$

uniformly over θ, which can be shown by standard arguments in deriving uniform convergence rate of kernel estimators with the aid of the Markov splitting techinique. Therefore, we have

$$Q_{n,3}(\theta) = O(\sqrt{T}h^2) + O_P(\sqrt{\Delta(\log n)/h}) \text{ uniformly over } \theta.$$

Therefore, we obtain $Q_n(\theta) = O_P(\sqrt{T}h^2 + \sqrt{\Delta(\log n)/h})$, which is $o_P(1)$ under the stated rate conditions on Δ and h. From these arguments, the asymptotic distribution of $\hat{Z}(\theta)$ is determined by $R_n(\theta)$. To investigate the limit behavior of $R_n(\theta)$, we note that it is the sum of a martingale difference array, to which we can apply the central limit theorem (CLT). In particular, we use Nishiyama's CLT (Sec. 2 of Nishiyama, 1996; Sec. 4 of Nishiyama, 2000), for which required conditions can be easily verified by using the uniform boundedness of the test function ϕ and its derivatives, Assumption 4), and the compactness assumption of the parameter space Θ. Now, the first assertion of the theorem follows. For verifying the second assertion, we consider an expansion of $\hat{\lambda}(\theta)$ by using the Ito formula, and then, we can show that the limit covariance kernel $\Lambda_0(\theta_1, \theta_2)|_{\theta=\theta_1=\theta_2}$ coincides with the limit of $\hat{\lambda}(\theta)$, $\lambda_0(\theta)$. The uniformity can be easily checked by the uniform boundedness of relevant functions and the compactness of Θ. The proof is completed. ■

A.2 Numerical integration

Here, we outline how to numerically implement integrations with respect to x and θ, to obtain the test statistic $\hat{J}$. First, observe that

$$\begin{aligned} &\int \frac{1}{h}K\left(\frac{X_{i\Delta} - x}{h}\right)\rho^{-1}(x)\, dx = 1 + \int \frac{1}{h}K\left(\frac{X_{i\Delta} - x}{h}\right)G(x)\, dx \\ &= 1 + [L_{i,h}(x)\, G(x)]_{-\infty}^{\infty} - \int L_{i,h}(x)\, g(x)\, dx = 2 - \int L_{i,h}(x)\, g(x)\, dx, \end{aligned}$$

where $L_{i,h}(x) := (1/h)\int_{-\infty}^{x} K((X_{i\Delta}-u)/h)\,du$; the second equality follows from the integration by parts; and the last equality holds since $L_{i,h}(\infty) = G(\infty) = 1$ and $L_{i,h}(-\infty) = G(-\infty) = 0$. Using this, we consider the following approximation:

$$\int \frac{1}{h} K\left(\frac{X_{i\Delta}-x}{h}\right) \rho^{-1}(x)\,dx \simeq 2 - \frac{1}{R}\sum\nolimits_{r=1}^{R} \frac{1}{h} K\left(\frac{X_{i\Delta}-x_r}{h}\right),$$

where $\{x_r\}_{r=1}^{R}$ is a computer-generated (pseudo) random sequence. As $\{x_r\}$, we in particular use a so-called low-discrepancy sequence based on the Halton sequence, i.e., we let $x_r = G^{\text{INV}}(a_r)$, where $\{a_r\}_{r=1}^{R}$ is the first R numbers of the base-2 Halton sequence on the unit interval $(0,1)$ and $G^{\text{INV}}(a) := \log(a/[1-a])$ (the inverse function of $G(x)$). By the integration method outlined here, we can obtain a numerical approximation $\hat{g}^{R}(\theta)$ to $\left\{\int_{\mathbb{R}} \hat{\Psi}(x;\theta)\,dx\right\}^2$ for each θ. To integrate $\hat{g}^{R}(\theta)$, we also consider the use of a Halton-based sequence $\{\theta_u\}_{u=1}^{U}$, where $\theta_u := 2b_u - 1$ and $\{b_u\}_{u=1}^{U}$ is the first U numbers of the base-3 Halton sequence on the unit interval $(0,1)$. Then, we have an approximation to the numerator of $\hat{J}$:

$$\int_{[-1,1]} \left\{\int_{\mathbb{R}} \hat{\Psi}(x;\theta)\,dx\right\}^2 d\theta \simeq \int_{[-1,1]} \hat{g}^{R}(\theta_u)\,d\theta \simeq \frac{1}{U}\sum\nolimits_{u=1}^{U} \hat{g}^{R}(\theta_u),$$

where we let $R = U = 100$ through our simulation study. By using the same method, we can also obtain an approximation to the denominator of $\hat{J}$, $\displaystyle\int_{[-1,1]} \hat{\lambda}(\theta)\,d\theta$.

References

[1] Aït-Sahalia, Y., J. Fan & J. Jiang (2010) Nonparametric tests of the Markov hypothesis in continuous-time models (2010), *The Annals of Statistics*, 38, 3129-3163.

[2] Amaro de Matos, J. & M. Fernandes (2007) Testing the Markov property with high frequency data, *Journal of Econometrics*, 141, 44-64.

[3] Andrews, D.W.K & W. Ploberger (1994) Optimal test when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.

[4] Applebaum, D. (2009) *Lévy Processes and Stochastic Calculus*, 2nd edition, Cambridge University Press.

[5] Bandi, F.M. & T.H. Nguyen (2003) On the functional estimation of jump-diffusion models, *Journal of Econometrics*, 116, 293-328.

[6] Bandi, F.M. & V. Corradi (2011) Nonparametric nonstationarity tests, Working paper, Johns Hopkins University and University of Warwick.

[7] Bandi, F. M. & Phillips, P.C.B. (2003) Fully nonparametric estimation of scholar diffusion models, *Econometrica*, 71, 241-283.

[8] Bibby, B.M. & M. Sørensen (1997) A hyperbolic diffusion model for stock prices, *Fianance and Stochastic*, 1, 25-41

[9] Berenguer-Rico, V. & J. Gonzalo (2011) Summability of stochastic processes, a generalization of integration and co-integration valid for non-linear processes. Working paper, Universidad Carlos III de Madrid.

[10] Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd edition, John Wiley and Sons.

[11] Boning, Wm. B. & Sowell F. (1999) Optimality for the integrated conditional moment test, *Econometric Theory*, 15, 710-718.

[12] Bierens, H.J. (1982) Consistent model specification tests, *Journal of Econometrics*, 20, 105-134.

[13] Bierens, H.J. (1990) A consistent conditional moment test of functional form, *Econometrica*, 58, 1443-1485.

[14] Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moments tests, *Econometrica*, 65, 1129-1151.

[15] Chen, B. & Y. Hong (2011) Testing for the Markov property in time series, forthcoming in *Econometric Theory*.

[16] De Jong, R.M. (1996) The Bierens test under data dependence, *Journal of Econometrics*, 72, 1-32.

[17] Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association, 74, 427-431.

[18] Dynkin, E.B. (1956) Infinitesimal operators of Markov processes, *Theory of Probability and its Applications* 1, 34-54.

[19] Dynkin, E.B. (1965) *Markov Processes*, Vol. I, Springer-Verlag.

[20] Escanciano, J. C. & D.T. Jacho-Chávez (2010 Approximating the critical values of Cramér-von Mises tests in general parametric conditional specifications, *Computational Statistics & Data Analysis*, 54, 625-636.

[21] Etheir, W. and T. G. Kurtz (1986) *Markov Processes*, Springer-Verlag.

[22] Granger, C.W.J. and N. Swanson (1997) An introduction to stochastic unit-root processes, *Journal of Econometrics* 80, 35-62.

[23] Hall, P. & Heyde, C.C. (1980) *Martingale Limit Theory and Its Application*, Academic Press.

[24] Hansen, B.E. (1996) "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica* 64, 413-430.

[25] Hansen, L.P. & Scheinkman, J.A. (1995) Back to the future: generating moment implications for continuous time Markov processes, *Econometrica*, 63, 767-804.

[26] Hansen, L.P., Scheinkman, J.A. and Touzi, N. (1998) Spectral Methods for Identifying Scalar Diffusions, *Journal of Econometrics*, 83, 1-32.

[27] Higham, D.J., X. Mao, & A.M. Stuart (2003) Strong convergence of Euler-type methods for nonlinear stochastic differential equations, *SIAM Journal on Numerical Analysis*, 40, 1041-1063.

[28] Höpfner, R. & E. Löcherbach (2003) Limit theorems for null recurrent Markov processes, *Memoirs of the American Mathematical Society*, 768.

[29] Jacob, N. (2002) *Pseudo Differential Operators and Markov Processes I*, Imperial College Press.

[30] Kallenberg, O. (2002) *Foundations of Modern Probability*, 2nd edition, Springer-Verlag.

[31] Karlin, S. & H.M. Taylor (1981) *A Second Course in Stochastic Processes*, Academic Process, New York.

[32] Karlsen, H.A. and Tjøstheim, D. (2001) Nonparametric Estimation in Null Recurrent Time Series, *The Annals of Statistics*, 29, 372-416.

[33] Kristensen, D. (2009) Uniform convergence rates of kernel estimators with heterogeneous dependent data, *Econometric Theory*, 25, 1433-1445.

[34] Komatsu, T. (1973) Markov processes associated with certain integro-differential operators, *Osaka Journal of Mathematics*, 10, 271-303.

[35] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, & Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, 54, 159-178.

[36] Nagakura, D. (2009) Asymptotic theory for explosive random coefficient autoregressive models and inconsistency of a unit root test against a stochastic unit root process, *Statistics and Probability Letters*, 79, 2476-2483.

[37] Nicolau, J. (2005) Processes with volatility-induced stationarity: An application for interest rates, *Statistica Neerlandica*, 59, 376-396.

[38] Nishiyama, Y. (1996) A central limit theorem for l^∞-valued martingale difference array and its application, Preprint 971, Department of of Mathematics, Utrecht University.

[39] Nishiyama, Y. (2000) Weak convegence of some classes of martingales with jumps, *The Annals of Probability*, 28, 685-712.

[40] Rogers, L.C.G., and Williams, D. (1994) *Diffusions, Markov Processes and Martingales*, Vol. 1, 2nd edition, Cambridge University Press.

[41] Schaumburg, E. (2004) Estimation of markov processes with Levy type generators," Working paper, Northwestern University.

[42] Shiryaev, A. N. (1989) *Probability*, 2nd edition, Springer-Verlag.

[43] Skorokhod, A.V. (1961) Stochastic equations for diffusion processes in a bounded region (I & II), *Theory of Probability and its Applications*, 6 & 7, 264-274 & 3-23.

[44] Stroock, D.W. (1975) Diffusion processes associated with Lévy generators, *Probability Theory and Related Fields*, 32, 209-244.

[45] Vanhems, A. (2006) Nonparametric study of solutions of differential equations, *Econometric Theory*, 22, 127-57.

[46] Wee, I.-S. (1999) Stability in multidimensional jump-diffusion processes, *Stochastic Processes and their Applications*, 80, 193-209.

[47] Wee, I.-S. (2000) Recurrence and transience for jump-diffusion processes, *Stochastic Analysis and Applications*, 18, 1055-1064.