Non-Parametric Specification Testing for Continuous-Time Markov Processes: Do the Processes Follow Diffusions?

Shin Kanaya*†

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

November 2008

Abstract

I propose a new non-parametric testing procedure to determine whether or not an underlying continuous-time process is a diffusion. While many papers in economics and finance presuppose that the dynamics of economic variables are described by diffusion processes, an empirical validation of the *diffusion hypothesis* is rarely found. I develop a new theorem which non-parametrically and fully identifies diffusion processes within a class of univariate time-homegeneous Markov processes through their infinitesimal generators - functional operators computed via derivatives of the conditional expectations with respect to time. I construct test statistics based on this theorem and derive their asymptotic distributions under the stationarity and mixing conditions. I also propose a simulation-based technique to approximate the asymptotic distributions, since the distributions of the original statistics depend upon a large number of unknown parameters

*E-mail address: shin.kanaya@economics.ox.ac.uk

†I am deeply indebted to Bruce E. Hansen for many helpful discussions, guidance and encouragement throughout this project. I am very grateful to Dennis Kristensen and Jack R. Porter for their valuable advice and suggestions. I also wish to thank James R. Walker, Tara M. Sinclair, Peter C. B. Phillips, Serena Ng, Phoebus J. Dhrymes, Alexei Onatski, Gautam Tripathi, Joon Park, Yoosoon Chang, Neil Shephard, Bent Nielsen, Andrew J. Pattton, Kevin Sheppard, Brendan K. Beare, Mika Meitz and seminar participants at University of Wisconsin-Madison, the 2007 Midwest Econometrics Group, Columbia University, Texas A&M, University of Connecticut and University of Oxford. Comments are welcome.

and functions. Monte-Carlo simulations are conducted to study the finite-sample size and power properties of the test. I apply the proposed method to short-term interest rates and foreign exchange rates to examine the validity of the diffusion hypothesis.

1 Introduction

Continuous-time diffusion models are widely used in economics and finance to describe the dynamics of economic variables such as asset prices, interest rates, and exchange rates. Unfortunately, economic theory puts only few restrictions on the specification of the processes, and as a consequence many parametric diffusion models coexist in the literature. To mitigate the risk caused by misspecification, several authors have proposed non-parametric methods to estimate diffusion processes. However, the fundamental question whether the process is actually a diffusion still remains. This paper is the first to provide a statistical method to directly answer this question.

The use of diffusions has been a classical starting point in building economic/financial models in the continuous-time framework, e.g., Merton (1973), Constantinides (1990) for stock prices; Vasicek (1977), Cox, Ingersoll and Ross (1985), Heath, Jarrow and Morton (1992) for interest rates; Froot and Obstfeld (1991), Krugman (1991) for exchange rates. It has been standard to assume that the underlying process follows a diffusion in the derivative security pricing literature since the seminal work of Black and Scholes (1973) and Merton (1974). The reason why diffusions are commonly and extensively used (as found in these examples) seems to be ascribable to their technical and theoretical aspects, not necessarily to their empirical aspects. They are analytically and computationally tractable: optimization theory and simulation/numerical techniques associated with diffusions are well-established. Diffusions are also tied to an important economic theory: they represent a class of processes which may not violate the no-arbitrage hypothesis if they are used to model the dynamics of price processes, since they constitute a subclass of semi-martingales. As is well known, if arbitrage opportunities are exhausted in a frictionless market environment, price processes must be semi-martingale under standard regularity conditions (see, e.g., Back (1991)). In fact, every semi-martingale process must be a diffusion under the path-continuity and Markov assumptions.[1]

While diffusions are appealing in these respects, their empirical plausibility has not been

[1] These assumptions are the sources of the tractability of diffusions.

fully checked, largely because there has been no statistical method to directly examine the diffusion hypothesis – the hypothesis that the process follows a diffusion. This paper is devoted to developing such a method. Our method is by necessity non-parametric since a class of diffusions cannot be characterized only by finite-dimensional parameters and alternatives to diffusions are generally unknown. Despite their flexibility in showing various distributional features (see, e.g., Stramer and Tweedie (1999) and Nicolau (2002, 2005)), diffusions fail to capture rare large moves, jumps, since their paths are by nature continuous. Accordingly, the alternatives include jump diffusions (e.g., see Merton (1976) and Johannes (2004)), and pure jump models (e.g., see Cox and Ross (1976), Barndorff-Nielsen and Shephard (2001)). Our method could be used to provide supporting evidence for picking diffusions among various alternative models.

I consider the case where the underlying stochastic process is within a class of continuous-time Markov processes, which of course includes diffusion processes. For each process in such a class, its corresponding *infinitesimal generator* is well-defined. The infinitesimal generator is a functional operator defined via a derivative of conditional expectations and can be used to derive implications/restrictions for Markov processes. Hansen and Scheinkman (1995) first employed infinitesimal generators in the econometrics literature for inference of parametric nonlinear Markov processes. They use infinitesimal generators to derive moment restrictions of stationary processes, and suggest GMM-based estimators and tests. This approach is also developed by Conley, Hansen, Luttmer and Scheinkman (1997), Zhou (2003), Duffie and Glynn (2004), Schaumburg (2004), Aït-Sahalia and Mykland (2007) and others.

In contrast, I use the infinitesimal generator to verify the type of process being observed. The infinitesimal generator takes a specific form depending on the type of process. For example, the generator for a diffusion process takes a linear parabolic differential form which consists of two terms. For a jump-diffusion process, the generator has one more term which represents the effect of jumps. While using transition probabilities is the most common and general way to define continuous-time Markov processes, there exists a one-to-one mapping between them and infinitesimal generators under some continuity conditions (which are stated in the next section). In this sense, infinitesimal generators can fully characterize Markov processes.[2] Based on this fact, we derive a new theorem which fully and non-parametrically

[2] There are at least four alternative ways to define/characterize Markov processes: (i) transition probabilities; (ii) conditional expectation operators (semi-group of conditional expectations); (iii) infinitesimal generators; (iv) stochastic differential equations. While (i) can define the widest class of process, (ii) can also define the same class under the continuity conditions on transition probabilities. Since there exists a

identifies diffusion processes within a class of univariate, time-homogeneous Markov processes by comparing infinitesimal generators. The theorem is useful in constructing a feasible testing procedure, which only requires us to check whether a certain moment is zero. We propose two estimators of the Nadaraya-Watson type for the infinitesimal generator: one estimator is consistent for general Markov processes; but the other is consistent only for diffusions. We construct test statistics based on this identification theorem and these two non-parametric estimators. Our testing procedure can be regarded as a variant of the so-called nuisance parameter (or Bierens') approach found in the conditional moment test literature (see Bierens (1982, 1990) and Stinchcombe and White (1998)). As is the case in this approach, the asymptotic distributions of the proposed statistics depend upon a large number of unknown parameters and functions, and thus tabulation of critical values is impossible. Therefore, we also propose a simulation-based technique to approximate the asymptotic distributions. We conduct Monte-Carlo simulations to check the finite-sample property of the testing procedure. According to them, the finite-sample performance of the (asymptotic and simulation-based) approximations seems satisfactory.

It is worth reviewing previous studies on specification testing for parametric diffusions developed by Aït-Sahalia (1996), Conley, Hansen, Lutter and Scheinkman (1997), Gao and King (2004), Hong and Li (2005), Aït-Sahalia, Fan and Peng (2006), and others. Their tests examine the null hypothesis that the process belongs to a certain family of parametrized diffusions. To see the contribution of my test, consider the case where the parametric null hypothesis is rejected in such specification tests. There are two possible reasons for the rejection: (i) the misspecification of the parametric restriction; (ii) the invalidity of the diffusion hypothesis. The true process may still be a diffusion, even when the rejection occurs as in (i). Unfortunately, these tests, which presuppose the diffusion hypothesis, cannot uncover the reason of the rejection. On the other hand, the proposed test here can determine the reason. If the rejection is due to (i), econometricians might be encouraged to explore more plausible parametric models. In this respect, the proposed test could extend the applicability of diffusions. As a practical procedure to pick a good parametric diffusion model, the following two-step procedure would be useful: **1.** examine the diffusion hypothesis using the proposed test; **2.** if the diffusion hypothesis is not rejected, try the parametric specification test with

one-to-one mapping between semi-groups of conditional expectations and the infinitesimal generators (shown by the Hille-Yoshida theorem), (i)-(iii) define the same class under the appropriate conditions (stated later). The last way by stochastic differential equations can probably describe a less wide class of processes than the others.

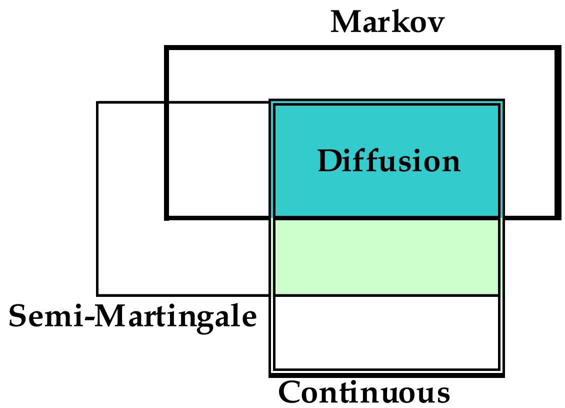

Figure 1: Relationship among four classes of processes

several parametric forms until a specific form is found that is not rejected.

Other studies have proposed testing procedures to see the presence of jumps in the paths of processes, e.g., Barndorff-Nielsen and Shephard (2006), Peters and de Vilder (2006), Andersen, Bollerslev, and Dobrev (2007), Aït-Sahalia and Jacod (2007). Their tests are, roughly speaking, based on the estimators of the quadratic variations of the process (= realized volatilities), and consequently examine the jump hypothesis within a class of (local) semi-martingales, for which quadratic variations are well-defined. Meanwhile, the diffusion process has continuous paths and any Markov process whose paths are continuous is, in fact, a diffusion. Therefore, the proposed test could be regarded as a test for the presence of jumps within a class of Markov processes. Figure 1, which presents the (rough) relationship among four relevant classes of processes, should be useful to see how my test is different from other jump tests.

The fact that diffusions represent a proper **sub**-class of continuous semi-martingales leads to the following two implications in our context. First, other jump tests can be interpreted as a conservative test if used to examine the diffusion hypothesis. If the no-jump hypothesis is rejected in such tests, then the diffusion hypothesis is also rejected. However, if it is not rejected, we cannot conclude the process follows a diffusion, and need additional evidence to obtain such a conclusion.[3] The proposed test, in contrast to this, could support the use of diffusions more directly. Second and more importantly, while other jump tests are basically designed for high-frequency data, my test does not necessarily require such data and works even for relatively low frequency data. In fact, our simulation study shows that the finite-

[3] Processes which are continuous semi-martingale but are not Markovian are, for example, given by solutions to stochastic differential equations with delay (see, e.g., Section 3.3 of Kutoyants (2004)).

sample performance of the test, based on asymptotic approximations, is satisfactory for daily observations. This finding is consistent with the arguments in Fan (2006a,b) and Phillips and Yu (2005, 2006).[4] It could also be explained as follows. Probabilistic structure implied by both the Markov and semi-martingale properties should be strong in comparison to that implied only by the latter property. My test could utilize such relatively strong structure to examine the null hypothesis. Other jump tests exploit only the semi-martingale property, which may be rephrased as the property of quadratic variations, and thus necessarily require high-frequency data.

An apparently possible approach to construct a testing procedure for Markov processes is to use transition densities/probabilities, as developed by Hong and Li (2005), Aït-Sahalia, Fan and Peng (2006), and others. However, for our purpose of non-parametrically identifying diffusions, it is difficult to take such an approach since the conditions on transition densities for the process to be a diffusion are generally unknown. Some of these restrictions are presented by Aït-Sahalia (2002a), Darolles, Florens and Gouriéroux (2004), but they are only either necessary or sufficient conditions, and far from full characterization. In contrast to this, the diffusion restriction can be easily and fully described by infinitesimal generators, and consequently we can develop a procedure to directly assess the diffusion hypothesis.

Since infinitesimal generators can fully characterize Markov processes, our approach can also be used to construct other types of testing procedures: e.g., specification testing for parametric or semi-parametric models, testing for jumps in non-parametric jump-diffusion models. We can develop any other test as long as we can estimate the infinitesimal generator which incorporates a certain restriction. Our infinitesimal-generator-based approach has several advantages. First, infinitesimal generators are simple to use. As discussed, one possible alternative is to use transition densities/probabilities. However, transition-based methods require us to compute model-implied transition densities, which is often a challenging task. For a general Markov process, the transition density rarely has a closed-form expression.[5] As a result, some technique to approximate the transition density is generally required, e.g., the simulation method of Pedersen (1995), Brandt and Santa-Clara (2002), the Hermite expansion method of Aït-Sahalia (1999, 2002b). In contrast, the infinitesimal generator is easy to estimate non-parametrically or parametrically, in particular when the process can be described

[4] They argue that biases in estimating the continuous-time diffusion which stem from the discretization of the process are less serious than other biases.

[5] The transition density (if it exists) is generally given as a solution to a particular partial differential equation, called the Kolmogorov equation (see, e.g., Lo (1988)).

by a certain stochastic differential equation. Second, our approach can extend a class of hypotheses which can be examined in a statistical testing framework. As is the case for my test of the diffusion hypothesis, many hypotheses have been hard to test, since it is not easy to find the restrictions (implied by the hypothesis) which allow us to develop a sensible test theory. Infinitesimal generators could provide convenient restrictions, in particular for hypotheses of the non-parametric and semi-parametric type.

The rest of the paper is organized as follows. The next section formally introduces infinitesimal generators for Markov processes with some examples. We also clarify technical requirements which should be imposed on the Markov processes we consider. Section 3 presents two nonparametric estimators for the infinitesimal generator which are used to construct test statistics. In section 4, the null and alternative hypotheses are presented formally, and a new theorem to non-parametrically identify the hypotheses is also presented. In section 5, I define the test statistics and investigate their asymptotic distributions. Sections 6 and 7 present simulation study results and empirical applications, respectively. Section 8 concludes. All proofs can be found in 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, respectively; and $f \cdot g(x)$ denotes a product of functions: $f(x)\,g(x)$. The symbols $\xrightarrow{p}$, $\xrightarrow{d}$ and $\Longrightarrow$ mean convergence in probability, in distribution, and weak convergence, respectively. The abbreviation a.s. means "almost surely."

2 Infinitesimal Generators for Markov Processes

Let $\{X_s\}_{s\geq 0}$ be a scalar, time-homogeneous and continuous-time Markov process defined on the filtered probability space $\left(\Omega, \mathfrak{F}, \{\mathfrak{F}_s\}_{s\geq 0}, \Pr\right)$ which satisfies the usual conditions. Denote by I and $\mathfrak{B}(I)$ the state space of $\{X_s\}$ and its Borel field ($\mathfrak{B}(I)$ is defined with respect to the Euclidean norm). We consider the case where $I = (-\infty, \infty)$, the whole real line.[6] The time-homogeneous Markov process is determined by the transition function: $P(s, x, \Gamma) :=$

[6] The other cases can be treated by considering some appropriate transformations by at least twice continuously differentiable functions. If $I = (0, \infty)$ for example, then we can consider the logarithmic transformation. Such a transformation does not change the property of the process we wish to examine: we can show by Ito's lemma that if the original process follows a diffusion, then the transformed process does also so. It is possible to develop a testing procedure for the case where $I \neq (-\infty, \infty)$, directly without relying on any transformation of the process. However, we do not pursue such an approach since it complicates our proofs.

$\Pr(X_s \in \Gamma | X_0 = x)$, expressing the probability that the process which has started from point x is in the set $\Gamma\,(\in \mathfrak{B}(I))$ at time s, and the initial distribution of X_0. Let π be the probability density of X_0, and assume that π is continuous on I. Time-homogeneity of the process means that $\Pr(X_s \in \Gamma | X_0 = x) = \Pr(X_{s+t} \in \Gamma | X_t = x)$ for any $t \geq 0$. Denote by $\mathbf{C}(I)$ the space of continuous functions from I to $\mathbb{R}$, and by $\hat{\mathbf{C}}(I) \subset \mathbf{C}(I)$ the space of continuous functions from I to $\mathbb{R}$ whose limits at infinity are zero, i.e. $\lim_{|x|\to\infty} f(x) = 0$, equipped with the sup-norm $||f|| := \sup_{x \in I} |f(x)|$. For each s, define the conditional expectation operator, a map from $\hat{\mathbf{C}}(I)$ into some functional space, as

$$\mathcal{T}_s \varphi(x) := E[\varphi(X_s) | X_0 = x] = \int \varphi(y)\, dP(s, x, dy) \quad \text{for} \quad \varphi \in \hat{\mathbf{C}}(I).$$

By the law of iterated expectations, $\{\mathcal{T}_s\}_{s\geq 0}$ satisfies a *semigroup* property, i.e., $\mathcal{T}_{s+t} = \mathcal{T}_s\mathcal{T}_t = \mathcal{T}_t\mathcal{T}_s$. We call $\{\mathcal{T}_s\}$ *a semigroup of the conditional expectations*, or simply *a semigroup.*[7] A semigroup $\{\mathcal{T}_s\}$ which satisfies the following conditions is called a *Feller* semigroup: (i) $\mathcal{T}_s : \hat{\mathbf{C}}(I) \to \mathbf{C}(I)$ for each s; (ii) $||\mathcal{T}_s\varphi - \varphi|| \to 0$ as $s \to 0$. A process whose conditional expectation operator satisfies (i)-(ii) is called a Feller process.[8] For the Feller semigroup associated with the process $\{X_s\}$, we define an *infinitesimal generator* $\mathcal{A} : \mathfrak{D}(\mathcal{A})\,(\subset \hat{\mathbf{C}}(I)) \to \mathbf{C}(I)$ as

$$\mathcal{A}\varphi(x) := \lim_{\Delta \to 0+} \frac{\mathcal{T}_\Delta \varphi(x) - \varphi(x)}{\Delta} = \lim_{\Delta \to 0+} \frac{E[\varphi(X_\Delta) | X_0 = x] - \varphi(x)}{\Delta}, \tag{1}$$

where $\mathfrak{D}(\mathcal{A})$ denotes the domain of $\mathcal{A}$, i.e., a family of functions in $\hat{\mathbf{C}}(I)$ for which the convergence of the right-hand side of (1) is with respect to the sup-norm $||\cdot||$. Note that $\mathcal{A}\varphi$ is necessarily continuous if it is well-defined, since $[\mathcal{T}_\Delta \varphi(x) - \varphi(x)]/\Delta$ is continuous in x for each Δ (by the Feller property (i)) and $\mathcal{A}\varphi(x)$ is its uniform limit.[9] We call a function φ in $\mathfrak{D}(\mathcal{A})$ a

[7] In this paper, we only consider a class of semigroups associated with time-homogeneous Markov processes. For general definitions of semigroups, see, e.g., Ethier and Kurtz (1986).

[8] Many other definitions of Feller processes can be found in the literature. One of them is based on a property of the corresponding transition functions, which is proved to be consistent with ours (see, e.g., Ch. 2 and 3 of Dynkin (1965)).

[9] While the sup-norm is used to define $\mathcal{A}$, there exists another way to define infinitesimal generators. For example, Hansen and Scheinkman (1995) define infinitesimal generators in the $L^2(Q)$ space, which is a Hilbert space and Q is an invariant (stationary) distribution of the process. In order to analyze some properties of Markov processes such as reversibility using inifinitesimal generators, we generally need to define them in a Hilbert space. However, for my purpose here, the definition in the Banach space $\mathbf{C}(I)$ is sufficient. Moreover, we heavily exploit the continuity of $\mathcal{A}\varphi$ to construct a feasible testing procedure, as will be seen in the following sections. If the generator is defined with respect to some integral-norm, e.g., the $L^2(Q)$-norm, such a continuity property will be lost. For these reasons, we work with the space $\mathbf{C}(I)$.

test function. 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 $s \geq 0$.

We restrict our attention to time-homogeneous Markov processes which are in a class of Feller processes.[10] Many processes, such as diffusions and more general Lévy-type processes, belong to a Feller class. In particular, most Markov processes found in economics and finance are actually proved to be Feller, and thus, our restriction to a Feller class does not exclude any interesting processes. This restriction can facilitate our analysis in two respects: first, the continuity of $\mathcal{A}\varphi$ is ensured for the Feller process, which allows us to use approximation theory (as we will see later). Second, infinitesimal generators fully characterize continuous-time Markov processes at least within a class of Feller processes in the following sense: the transition function is uniquely determined by the corresponding generator (for details, see Dynkin (1956))[11] and thus, there exists a one-to-one mapping between infinitesimal generators and continuous-time Markov processes. This full characterization property allows us to construct a consistent testing procedure within a Feller class. Throughout we work under the condition that $\{X_s\}$ is a time-homogeneous Feller process and that its corresponding infinitesimal generator $\mathcal{A}$ has a non-empty domain $\mathfrak{D}(\mathcal{A})$.

The form of the generator depends on the type of the process. Consider $\{X_s\}_{s\geq 0}$ which is represented by the following stochastic differential equation (SDE):

$$dX_s = \mu(X_s)\,ds + \sigma(X_s)\,dW_s \tag{2}$$

where $\mu(\cdot)$ is the drift function; $\sigma^2(\cdot) > 0$ is the diffusion (volatility) function; and $\{W_s\}_{s\geq 0}$ is a standard Brownian motion. The infinitesimal generator $\mathcal{A}$ associated with this process is given by

$$\mathcal{A}\varphi(x) = \mathcal{A}_0\varphi(x) := \mu(x)\,\varphi'(x) + \frac{1}{2}\sigma^2(x)\,\varphi''(x) \tag{3}$$

for a test function φ which is at least twice continuously differentiable. Here, we formally define *a diffusion process* as a Feller process which is homogeneous in time and whose infinitesimal generator takes the form of (3).[12] This definition guarantees that any diffusion

[10] Note the following facts: (i) any Feller process has a modification whose sample path is right continuous; (ii) any Feller process with such a path is also a Markov process (see, e.g., Dynkin (1956) or Kallenberg (2002)). By (i) and (ii), Feller processes represent a subclass of Markov processes.

[11] More accurately speaking, if the infinitesimal generators of two Markov processes are equal, then the transition functions are also equal. For a general class of Markov processes, we only have the weaker assertion that the infinitesimal generator determines the finite-dimensional distributions of the process (refer to arguments on the Hill-Yoshida theorem and Proposition 4.1.6 in Ethier and Kurtz (1986)).

[12] Some authors may give a different definition from mine. Mine is a narrow definition, but is not inconsistent

process can be represented by the stochastic differential equation (2).[13] We can also check that any process described by the stochastic differential equation (2) is a Feller process and its generator takes the form of (3) (the former is proved in Kallenberg (2002) and the latter can be checked by Ito's lemma). Thus, we can define diffusion processes either by (2) or by (3).

Our basic strategy to construct a test for the diffusion hypothesis consists of two steps: we first estimate the infinitesimal generator non-parametrically in two ways based on (1) and (3), and second we compare the two estimates. As we will see later, the estimator based on (1) is consistent for general Markov processes, but the one based on (3) is only consistent for diffusions. Therefore we can conclude that the diffusion hypothesis is valid if the two estimates are close to each other, but invalid otherwise. In the next two sections we present two estimators for the infinitesimal generator and develop a method to compare the two estimated generators. Our main focus is on processes of dimension one, but some multivariate processes are also in our scope, as discussed later in this section.

To achieve existence and uniqueness of a solution to (2), we generally need to impose some conditions on μ and σ^2, e.g., boundedness of the growth rate and/or the smoothness. While the existence of a unique strong solution to (2) is often required in some specific applications, the existence of a weak solution should be sufficient for our econometric purpose. The existence of the weak solution is ensured under fairly weak conditions (various conditions can be found in Ch. 5.5 of Karatzas and Shreve (1991) and Ch. V of Protter (2005)). We do not pursue such conditions in this paper, since we only require that the process has a representation by the SDE (2) if it is a diffusion (uniqueness is not necessarily required). Instead, we impose some conditions on the domain of the generator in the next section, which may restricts the functional forms of the drift and diffusion functions to some extent.

Other examples of infinitesimal generators (of Feller processes) are listed below:

Markov Jump Process A Markov jump process $\{X_s\}$ is described by the following two components. Let $P(x, \Gamma)$ be a transition function on $I \times \mathfrak{B}(I)$, and $\lambda : x \to \mathbb{R}^+$ be a bounded continuous function. Timings when changes (jumps) in states occur are determined by a Poisson process with intensity parameter $\lambda(x)$ if the current state is x. If a jump occurs, then the transition probability is given by $P(x, \cdot)$. The infinitesimal generator of this process

with Ikeda and Watanabe (1981).

[13] This assertion can be proved by arguments on properties of Feller processes and the so-called martingale problem (see Ch. 4 & 5 in Ethier and Kurtz (1984) and Ch. IV of Ikeda and Watanabe (1981) for the details.

is characterized by

$$\mathcal{A}\varphi(x) = \lambda(x) \int [\varphi(y) - \varphi(x)] P(x, dy). \tag{4}$$

For more details of Markov jump processes, see pp.162-163 in Ethier and Kurtz (1986) or Hansen and Scheinkman (1995).

Jump-Diffusion Process A (time-homogeneous) jump-diffusion process is described by the following type of stochastic differential equation

$$dX_s = \mu\left(X_{s_-}\right) ds + \sigma\left(X_{s_-}\right) dW_s + dJ_s \tag{5}$$

where $dJ_s \left(= X_s - X_{s_-}\right)$ is a pure jump component of the process written as

$$dJ_s = \int_{\mathbb{R}\setminus\{0\}} g(X_{s-}, y) N(ds, dy).$$

N is a Poisson random measure having finite intensity measure $v(A) = E[N(1, A)]$ for any Borel set $A (\subset \mathbb{R}\setminus\{0\})$ and $g(x, y)$ stands for the impact of a jump conditional on the current state x where y is a random variable with distribution q. Non-parametric estimation methods for jump-diffusions are developed by Bandi and Nguyen (2003), Johannes (2004) and others. The generator for this type of process is given by

$$\mathcal{A}\varphi(x) = a(x)\varphi'(x) + \frac{1}{2}b^2(x)\varphi''(x) + \lambda(x)\int_{\mathbb{R}\setminus\{0\}} [\varphi(x + g(x, y)) - \varphi(x)] q(dy) \tag{6}$$

where λ is defined via an argument on the compensation:

$$\begin{aligned} E\left[\int_0^t \int g\left(X_{s_-}, y\right) N(ds, dy) \middle| X_{s-} = x\right] &= \lambda(x) E[g(X_{s-}, y) | X_{s-} = x] t \\ &= \lambda(x) \int_{\mathbb{R}\setminus\{0\}} g(x, y) q(dy) t. \end{aligned}$$

General Lévy-type Process All processes presented above are included in a class of more general processes, Lévy-type processes, whose corresponding infinitesimal generators are represented as

$$\begin{aligned} \mathcal{A}\varphi(x) = {} & \mu(x)\varphi'(x) + \frac{1}{2}\sigma^2(x)\varphi''(x) \\ & + \int_{\mathbb{R}\setminus\{0\}} \left[\varphi(x+z) - \varphi(x) - \frac{z}{1+z^2}\varphi'(z)\right] \nu(dz; x) \end{aligned} \tag{7}$$

where ν is the Lévy kernel satisfying

$$\int_{\mathbb{R}\setminus\{0\}} \frac{z^2}{1+z^2} v(dz, x) < \infty. \tag{8}$$

$\nu(dz, x)$ is interpreted as the expected number of jumps whose size is in the (small) interval "dz" per unit of time, conditional on the current state x. The only restriction on ν is (8), and it allows $\int_{|y|>0} v(dz, x) = \infty$ which corresponds to an infinite number of jumps within a finite period of time. We can confirm that (3), (4) and (6) are actually represented as (7) by suitable reparametrization.[14] For a general reference of Lévy(-type) processes, see Bertoin (1998) or Applebaum (2004).

Stochastic Volatility Model Another interesting example is a class of stochastic-volatility (SV) processes/models. One such model is given by the following stochastic differential equations:

$$\begin{cases} dX_s = \mu(X_s)\, ds + \sqrt{V_s} dW_{1s}, \\ dV_s = b(V_s)\, ds + a(V_s)\, dW_{2s} \end{cases} \tag{9}$$

where $\{X_s\}$ is an observable process, V_s is an unobservable volatility process, and $\{W_{1s}\}$ and $\{W_{2s}\}$ are standard Brownian motions (possibly correlated). Unfortunately, it is not possible to distinguish this type of process from (2) in our framework for univariate processes. To see this point, compute the right-hand side of (1) for the SV model. Using Ito's lemma, we have

$$\lim_{\Delta \to 0+} \{E[\varphi(X_\Delta) | X_0 = x] - \varphi(x)\} / \Delta = \mu(x) \varphi'(x) + \frac{1}{2} E[V_0 | X_0 = x] \varphi''(x). \tag{10}$$

This is the linear second-order differential operator since $E[V_s | X_s = x]$ is a function of x, and thus takes the same form as $\mathcal{A}_0 \varphi(x)$ in (3). The infinitesimal generator corresponding to (9) should be defined on the space of functions $I \times I \to \mathbb{R}$. Unobservability of the latent volatility process $\{V_s\}$ hampers the estimation of its infinitesimal generator. Since the SV model as in (9) should actually be called a diffusion "of dimension two," there is no surprise that (2) and (9) are not distinguished only by looking at a single dimensional aspect of the generators. In general, the whole system of multi-dimensional processes needs to be observable (or at least estimable) for constructing a testing procedure based on its infinitesimal generator.[15] Analyzing multi-dimensional processes with infinitesimal generators is also possible, but it is beyond the scope of this paper – we need not to consider the multi-dimentional case to examine the diffusion hypothesis. The multi-dimensional extension will be left for future work.

[14] The notations in (7) and (8) follow those in Schaumburg (2004).

[15] Kristensen (2006) and Kanaya and Kristensen (2007) propose nonparametric estimators for the latent volatility process $\{V_s\}$, $b(\cdot)$ and $a(\cdot)$ for SV models of a particular type as in (9), and open up the possibility to distinguish them from one-dimentional diffusions as in (2).

Pure Jump JDSV Model While the "traditional" stochastic volatility model (10) cannot be distinguished from a diffusion class without the knowledge of an unobservable component, a class of stochastic volatility models driven solely by jumps can. Barndorff-Nielsen and Shepahrd (2001), Carrm, Geman, Madan and Yor (2003), Todorov (2006) and others introduce such models. One such model is a pure jump Jump-Driven Stochastic Volatility (JDSV) model, shown in Todorov (2006), written as

$$\begin{cases} dX_s = \int_{\mathbb{R}\setminus\{0\}} \sigma_{s-} g(y)\, d\tilde{N}(ds, dy), \\ d\sigma_s^2 = \int_{\mathbb{R}\setminus\{0\}} k(y)\, dN(ds, dy) \end{cases} \tag{11}$$

where N is a Poisson random measure, $\tilde{N}$ is its compensated version, $g : \mathbb{R}\setminus\{0\} \to \mathbb{R}\setminus\{0\}$, and $k : \mathbb{R}\setminus\{0\} \to \mathbb{R}^+$. For this model, (1) can be computed as

$$\begin{aligned} &\lim_{\Delta \to 0+} \{E[\varphi(X_\Delta) | X_0 = x] - \varphi(x)\} / \Delta \\ &= \alpha(x) \varphi'(x) + \int_{\mathbb{R}\setminus\{0\}} [\varphi(x + f(x) g(y)) - \varphi(x)]\, \nu(dy; x) \end{aligned} \tag{12}$$

where α and f are some suitable functions, and ν is some suitable measure. Clearly, (11) is distinguished from (2).[16]

3 Non-Parametric Estimation of the Infinitesimal Generator

In this section, we consider two Nadaraya-Watson type estimators for the infinitesimal generators (1) (evaluated at some test function φ). Our test statistics are constructed based on these two estimators. Assume we have a dataset $\{X_{i\Delta}\}_{i=1}^n$ where n is the number of observations, and Δ is the time distance between adjacent observations. The time span of the data is $T = \Delta n$. Our asymptotic scheme consists of two assumptions: infill $(\Delta \to 0)$ and long-span $(T \to \infty)$. These two assumptions also imply the large-sample asymptotics of $n (= T/\Delta) \to \infty$. As is argued in Bandi and Phillips (2003), these two assumptions are necessary to estimate continuous-time (diffusion) process fully non-parametrically (see also Merton (1980)).

[16] The process represented by (11) may not be stationary and thus beyond the scope of this paper, although some suitable modifications ensure the stationarity.

We first consider the estimator which should be consistent for a wide class of Markov (Feller) processes: we estimate $\mathcal{A}\varphi(x)$ by

$$\widehat{\mathcal{A}\varphi}(x) = \frac{(1/Th)\sum_{j=1}^{n-1} K\left(\frac{X_{j\Delta}-x}{h}\right)\left[\varphi\left(X_{(j+1)\Delta}\right)-\varphi\left(X_{j\Delta}\right)\right]}{(\Delta/Th)\sum_{j=1}^{n} K\left(\frac{X_{j\Delta}-x}{h}\right)} \equiv \frac{\widehat{\mathcal{A}\varphi\cdot\pi}(x)}{\hat{\pi}(x)} \tag{13}$$

where K is a conventional kernel function and h is the bandwidth, $\widehat{\mathcal{A}\varphi\cdot\pi}(x)$ is the estimator of $\mathcal{A}\varphi(x)\times\pi(x)$, and $\hat{\pi}(x)$ is the estimator of the invariant density $\pi(x)$ (as we will see later, we assume that π is also the invariant (stationary) density of the process).

The consistency of this estimator for general Feller processes could be verified by using the two properties of the Feller semigroup (i) and (ii) in the previous section. For processes which are represented by some stochastic differential equations, verifying consistency should follow from arguments similar to those in Bandi and Phillips (2003) and Bandi and Nguyen (2003).

We next consider the estimator which incorporates the restriction that the process is a diffusion. This estimator is based on the infinitesimal first and second conditional moments:

$$\mathbf{M}_1(x) := \lim_{\Delta\to 0+}\frac{1}{\Delta}E\left[X_\Delta - x|X_0 = x\right]; \tag{14}$$

$$\mathbf{M}_2(x) := \lim_{\Delta\to 0+}\frac{1}{\Delta}E\left[(X_\Delta - x)^2|X_0 = x\right]. \tag{15}$$

If the process is a diffusion, then $\mathbf{M}_1(x) = \mu(x)$ and $\mathbf{M}_2(x) = \sigma^2(x)$, where μ and σ are the coefficients in the SDE (2).[17] While the infinitesimal first and second conditional moments are always well-defined for diffusion processes in the sense stated in the previous footnote, for

[17] While (14) and (15) are the usual definitions for the drift and diffusion functions for diffusion processes, $\mathbf{M}_2$ (and $\mathbf{M}_1$) may not be well-defined for some diffusion processes, since the relevant (conditional) moments on the right-hand sides of (14) and (15) do not necessarily exist. Even in such cases, both the drift and diffusion functions can be defined by modifying $\mathbf{M}_1$ and $\mathbf{M}_2$ in the following manner: for arbitrary $\varepsilon\,(>0)$

$$\mathbf{M}_{\varepsilon,1}(x) := \lim_{\Delta\to 0+}\frac{1}{\Delta}E\left[(X_\Delta - x)\,\mathbf{1}_{\{|X_\Delta - x|\le\varepsilon\}}|X_0 = x\right];$$

$$\mathbf{M}_{\varepsilon,2}(x) := \lim_{\Delta\to 0+}\frac{1}{\Delta}E\left[(X_\Delta - x)^2\,\mathbf{1}_{\{|X_\Delta - x|\le\varepsilon\}}|X_0 = x\right].$$

The truncated versions $\mathbf{M}_{\varepsilon,1}$ and $\mathbf{M}_{\varepsilon,2}$ are *always* well-defined for any diffusion, and coincide with μ and σ^2 in the SDE, respectively, for any $\varepsilon > 0$. Furthermore, if $\mathbf{M}_i$ is well-defined, $\mathbf{M}_i = \mathbf{M}_{\varepsilon,i}$ for any $\varepsilon > 0$. These results can be shown by exploiting the continuity property of the transition functions of diffusion processes (see, e.g., Section 5.4 of Friedman (1975)). For notational simplicity, we identify $\mathbf{M}_i$ with $\mathbf{M}_{\varepsilon,i}$ in what follows. If the former is not well-defined, it should be interpreted as the latter.

our testing purposes, they also need to be well-defined for the other type of process. Thus, we impose the following assumption on the underlying process $\{X_s\}$:

Assumption 1 *The right-hand sides of (14) and (15) are well-defined for each* $x \in I$ *and* $\mathbf{M}_1$ *and* $\mathbf{M}_2$ *are measurable functions on* I.

Note that we do not require that $\mathbf{M}_i$ is defined in the same manner as $\mathcal{A}\varphi$ is defined in Section 2 (i.e. we require pointwise convergence but not uniform convergence). The conditions in Assumption 1 are mild and are satisfied for virtually all Feller processes (e.g. (4), (6) and almost all Lévy-type processes). Explicit expressions of $\mathbf{M}_1(x)$ and $\mathbf{M}_2(x)$ for jump-diffusion processes are given in Bandi and Nguyen (2003). By (3) and the arguments above, the infinitesimal generator of the diffusion process should be written as

$$\mathcal{A}_0\varphi(x) = \mathbf{M}_1(x)\,\varphi'(x) + \frac{1}{2}\mathbf{M}_2(x)\,\varphi''(x), \tag{16}$$

where $\mathcal{A}_0$ does not generally coincide with $\mathcal{A}$. We define $\mathcal{A}_0$ by (16) for a general Feller process. We call $\mathcal{A}_0$ an *infinitesimal diffusion operator* of the process, since it is a differential operator which incorporates the diffusion restriction.

We estimate $\mathcal{A}_0\varphi(x)$ by

$$\widehat{\mathcal{A}_0\varphi}(x) = \widehat{\mu\cdot\varphi'}(x) + \frac{1}{2}\widehat{\sigma^2\cdot\varphi''}(x) \quad \text{for each } x \tag{17}$$

where $\widehat{\mu\cdot\varphi'}(x)$ and $\widehat{\sigma^2\cdot\varphi''}(x)$ are kernel-based nonparametric estimators (of the Nadaraya-Watson type) for $\mu\cdot\varphi'(x)$ and $\sigma^2\cdot\varphi'(x)$, respectively:

$$\widehat{\mu\cdot\varphi'}(x) = \frac{(1/Th)\sum_{j=1}^{n-1} K\left(\frac{X_{j\Delta}-x}{h}\right)\varphi'(X_{j\Delta})\left[X_{(j+1)\Delta}-X_{i\Delta}\right]}{(1/Th)\,\Delta\sum_{j=1}^{n} K\left(\frac{X_{j\Delta}-x}{h}\right)} \equiv \frac{\widehat{\mu\cdot\varphi'\cdot\pi}(x)}{\hat{\pi}(x)}; \tag{18}$$

$$\widehat{\sigma^2\cdot\varphi''}(x) = \frac{(1/Th)\sum_{j=1}^{n-1} \left(\frac{X_{j\Delta}-x}{h}\right)\varphi''(X_{j\Delta})\left[X_{(j+1)\Delta}-X_{i\Delta}\right]^2}{(1/Th)\,\Delta\sum_{j=1}^{n} K\left(\frac{X_{j\Delta}-x}{h}\right)} \equiv \frac{\widehat{\sigma^2\cdot\varphi''\cdot\pi}(x)}{\hat{\pi}(x)}. \tag{19}$$

$\widehat{\mu\cdot\varphi'\cdot\pi}(x)$ and $\widehat{\sigma^2\cdot\varphi''\cdot\pi}$ are the estimators for their corresponding objects.[18] These estimators are proved to be (uniformly) consistent for diffusion processes, even in the case that only $\mathbf{M}_{\varepsilon,i}(x)$ is well-defined and $\mathbf{M}_i(x)$ is not, as long as a test function is suitably chosen (see Kanaya (2007)).

[18] Some other estimators of $\mu(x)\,\varphi'(x)$ and $\sigma^2(x)\,\varphi''(x)$ could be used: for example, $\hat{\mu}(x)\,\varphi'(x)$ and $\hat{\sigma}^2(x)\,\varphi''(x)$ where $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ are the usual estimators of the drift and the diffusion functions found in

We conclude this section by listing several properties that the kernel function $K(\cdot)$ in (13), (18) and (19) should satisfy

Assumption 2 *(i) $K(\cdot)\,(:\mathbb{R}\to\mathbb{R})$ is twice continuously differentiable;*

(ii) $\int_{-\infty}^{\infty} K(x)\,dx = 1$ and $\int_{-\infty}^{\infty} xK(x)\,dx = 0$;

(iii) there exists some $\bar{K}\in(0,\infty)$ such that $\sup_{x\in\mathbb{R}}\left|K^{(k)}(x)\right|\le\bar{K}$, $\int\left|K^{(k)}(x)\right|dx\le\bar{K}$ and $\int_{-\infty}^{\infty} x^2\left|K^{(k)}(x)\right|dx\le\bar{K}$ for $k=0,1,2$.

4 The Null Hypothesis and its Identification

In this section, we present formally the null hypothesis and derive its identification theorem. The latter is required to construct feasible test statistics. Under the assumption that $\{X_s\}$ is a Feller process, we wish to examine the null hypothesis that the process is a diffusion. By the discussions in the previous sections, this null hypothesis is equivalent to

$$H_0:\mathcal{A}\varphi(\cdot)=\mathcal{A}_0\varphi(\cdot)\quad\text{for all }\varphi\in\mathfrak{D}(\mathcal{A})\cap\mathfrak{D}(\mathcal{A}_0)\tag{22}$$

where $\mathcal{A}$ and $\mathcal{A}_0$ are respectively defined by (1) and (16); $\mathfrak{D}(\mathcal{A})$ and $\mathfrak{D}(\mathcal{A}_0)$ are the corresponding domains. The alternative hypothesis H_1 is

$$H_1:\mathcal{A}\varphi(\cdot)\neq\mathcal{A}_0\varphi(\cdot)\quad\text{for some }\varphi\in\mathfrak{D}(\mathcal{A})\cap\mathfrak{D}(\mathcal{A}_0).$$

The stated null hypothesis requires us to check the equality in (22) for every test function in the intersection of the domains, which in turn requires knowledge of the domains $\mathfrak{D}(\mathcal{A})$ and $\mathfrak{D}(\mathcal{A}_0)$.[19] $\mathfrak{D}(\mathcal{A}_0)$ should consist of twice continuously differentiable functions in light of its

the literature:

$$\hat{\mu}(x)=\sum_{j=1}^{n-1}K((X_{j\Delta}-x)/h)\left[X_{(j+1)\Delta}-X_{i\Delta}\right]\Big/\Delta\sum_{j=1}^{n-1}K((X_{j\Delta}-x)/h);\tag{20}$$

$$\hat{\sigma}^2(x)=\sum_{j=1}^{n-1}K((X_{j\Delta}-x)/h)\left[X_{(j+1)\Delta}-X_{i\Delta}\right]^2\Big/\Delta\sum_{j=1}^{n-1}K((X_{j\Delta}-x)/h),\tag{21}$$

or estimators developed in Bandi and Phillips (2003) or Xu (2006). However, the asymptotic analysis with (18) and (19) is much easier and done under weaker regularity conditions than that with (20) and (21).

[19] H_0 and H_1 only concern test functions which are in the union of $\mathfrak{D}(\mathcal{A})$ and $\mathfrak{D}(\mathcal{A}_0)$. We may need to consider the case where $\mathcal{A}\varphi=\mathcal{A}_0\varphi$ for any $\varphi\in\mathfrak{D}(\mathcal{A})\cap\mathfrak{D}(\mathcal{A}_0)$, but there are some $\varphi\in\mathfrak{D}(\mathcal{A})$ such that $\mathcal{A}_0\varphi$ is not defined (i.e. $\varphi\notin\mathfrak{D}(\mathcal{A}_0)$). However, in this case, $\mathcal{A}$ should be regarded as an extension of $\mathcal{A}_0$ on the wider domain $\mathfrak{D}(\mathcal{A})\cup\mathfrak{D}(\mathcal{A}_0)$, where the existence of such an extended operator is confirmed by means of the Hahn-Banach theorem ($\mathcal{A}$ itself can be taken as an dominant operator of $\mathcal{A}_0$). Thus $\mathcal{A}$ and $\mathcal{A}_0$ should be identified on the wider domain, and therefore it is enough to compare the generators only on $\mathfrak{D}(\mathcal{A})\cap\mathfrak{D}(\mathcal{A}_0)$ (the same argument is applied to the case where $\mathcal{A}\varphi=\mathcal{A}_0\varphi$ for any $\varphi\in\mathfrak{D}(\mathcal{A})\cap\mathfrak{D}(\mathcal{A}_0)$ but for some $\varphi\in\mathfrak{D}(\mathcal{A}_0)$, $\mathcal{A}\varphi$ is not defined).

definition in (16), but $\mathfrak{D}(\mathcal{A})$ is hard to characterize. In general, it is not easy to characterize the domain of the infinitesimal generator,[20] and it is actually impossible here since we do not know the type of process $\{X_s\}$. One way to overcome this difficulty is to restrict the domain of the generator to some smaller set of test functions which is common for many classes of generators. However, such a restriction should be imposed carefully since it may result in losing information and yield lower power of the corresponding test. Fortunately, we can construct such a reduced class of functions without any information loss. Our approach to reduce functions is based on the concept of a core and "approximation" theory.

We first explain a core. We call $\mathfrak{C}(\mathcal{A})\,(\subset \mathfrak{D}(\mathcal{A}))$ a *core* of the generator $\mathcal{A}$ (or of the process which induces $\mathcal{A}$), if for each $\varphi \in \mathfrak{D}(\mathcal{A})$ there exists a sequence $\{\varphi_l : l = 1, 2 \ldots\}$ ($\varphi_l \in \mathfrak{C}(\mathcal{A})$) such that $\lim_{l\to\infty} ||\varphi_l - \varphi|| = 0$ and $\lim_{l\to\infty} ||\mathcal{A}\varphi_l - \mathcal{A}\varphi|| = 0$.[21] Denote by $\mathbf{C}_K^k(I)$ the set of all k-times differentiable functions (on I) with compact supports. For many Feller processes (and thus diffusion processes), $\mathbf{C}_K^\infty(I)$ is known to serve as a core (see, e.g., Ch. 8 of Ethier and Kurtz (1986)). It also seems standard to suppose that $\mathbf{C}_K^\infty(I)$ is a core in analyzing Feller processes (see, e.g., Section 4.5 of Jacob (2001); Section 2.6 Jacob (2002)). The following theorem shows that checking the difference of $\mathcal{A}\varphi$ and $\mathcal{A}_0\varphi$ for every $\varphi \in \mathfrak{D}(\mathcal{A}) \cap \mathfrak{D}(\mathcal{A}_0)$ is actually equivalent to checking it for every function in the reduced class $\mathfrak{C}$ of functions, under the milder condition for the existence of a core.

Theorem 1 *Suppose that Assumption 1 holds; and that there exists a non empty core* $\mathfrak{C}$ *of* $\mathcal{A}$ *such that* $\mathfrak{C} \subset \mathbf{C}_K^2(I)$. *Then,* $\forall \varphi \in \mathfrak{C}$, $||\mathcal{A}\varphi - \mathcal{A}_0\varphi|| = 0$ *if and only if*

$$\forall \varphi \in \mathfrak{D}(\mathcal{A}) \cap \mathfrak{D}(\mathcal{A}_0), \quad ||\mathcal{A}\varphi - \mathcal{A}_0\varphi|| = 0. \tag{23}$$

Note that $\mathbf{C}_K^2(I) \subset \mathfrak{D}(\mathcal{A}_0)$, and thus any element in $\mathfrak{C}$ is also a member of $\mathfrak{D}(\mathcal{A}_0)$. To exploit this result, we assume that $\{X_s\}$ has a non-empty core in the space of (at least) twice continuously differentiable functions. While this theorem claims that it suffices to look at test functions in the core $\mathfrak{C}$ to detect the difference of two generators $\mathcal{A}$ and $\mathcal{A}_0$, $\mathfrak{C}$ is still large and computing $\mathcal{A}\varphi$ and $\mathcal{A}_0\varphi$ for every $\varphi \in \mathfrak{C}$ is not an easy task. Therefore, we consider a further reduction of functions, where we borrow a result from approximation theory. The

[20] For the diffusion process, Hansen, Scheinkman and Touzi (1998) derive a full characterization of the domain of the infinitesimal generator which is defined with respect to the $L^2(Q)$ norm where Q is the invariant density of the process. Unfortunately, their result is not immediately applicable to our context since our generator is defined with respect to the sup-norm.

[21] In other words, $\mathfrak{C}(\mathcal{A})$ is a core if the closure of $\{(\varphi, \mathcal{A}\varphi) : \varphi \in \mathfrak{C}(\mathcal{A})\}$ is equal to $\{(\varphi, \mathcal{A}\varphi) : \varphi \in \mathfrak{D}(\mathcal{A})\}$,or if the former is dense in the latter.

next lemma states that any function φ in $\mathbf{C}_K^k(I)$ can be well-approximated by a sequence of weighted polynomials. Below, we denote by $\mathfrak{w}$ a weighting function:

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

Lemma 1 *For any function $\varphi \in \mathbf{C}_K^k(\mathbb{R})$, there exists a sequence of functions $\{H_J(x) : J = k+2, k+3, \dots\}$ such that*

$$\sum_{i=0}^{k-1} \sup_{x\in\mathbb{R}} \left|\varphi^{(i)}(x) - H_J^{(i)}(x)\right| \to 0 \quad \text{as} \quad J \to \infty \tag{24}$$

where $H_J(x) = L_J(x)\,\mathfrak{w}(x)$ and $L_J(x)$ is a polynomial function of degree at most $J-1$.

Theorem 1 and this lemma lead to useful theorems for identifying the null and alternative hypotheses under the following conditions:

Assumption 3 *(i) $H_J(x)$ is in the domain of $\mathcal{A}$ for any J.*
(ii) There exist some $c \in [0, 1/2)$ and $\bar{k}\,(\geq 2)$ such that if $\{\varphi_J : J = 1, 2, \dots\}$ is a sequence of functions in $\mathfrak{D}(\mathcal{A})$ approximating $\varphi\,(\in \mathfrak{D}(\mathcal{A}))$ in the sense that

$$\sum_{i=0}^{\bar{k}} \sup_{x\in I} \left|\varphi^{(i)}(x) - \varphi_J^{(i)}(x)\right| \to 0 \quad (\text{as } J \to \infty), \tag{25}$$

then,

$$\sup_{x\in I} \left|\{\mathcal{A}\varphi(x) - \mathcal{A}\varphi_J(x)\}\,w_c(x)\right| \to 0 \quad (\text{as } J \to \infty). \tag{26}$$

where $w_c(x) = \exp\{-cx^2\}/\sqrt{2\pi}$.
(iii) There exists a core $\mathfrak{C}(\subset \mathfrak{D}(\mathcal{A}) \subset \hat{\mathbf{C}}(I))$ of $\mathcal{A}$ of which each element is in $\mathbf{C}_K^{\bar{k}+1}(I)$.

Remark 1 *Condition (i) of Assumption 3 is required to approximate functions in the core by weighted polynomials since we use the result of Lemma 1. This condition is satisfied if $x^l\mathfrak{w}(x)$ is in the domain of $\mathcal{A}$ for each l. For diffusion processes (i.e., $\mathcal{A} = \mathcal{A}_0$), condition (ii) is satisfied under the following weak condition that the drift and diffusion functions are uniformly bounded by some polynomial functions.*[22] *For more general Lévy-type processes whose infinitesimal generator is characterized by (7), Assumption 3 is also satisfied when $\alpha(x)$, $\beta(x)$ and $\int_{\mathbb{R}\setminus\{0\}} \nu(dy; x)$ are bounded by some polynomial functions. For these two cases, $\bar{k} = 2$ in (25) is sufficient for (26) to hold.*

[22] Some parametric diffusion models found in the literature do not satisfy this condition. However, by modifying weighting functions $\mathfrak{w}$ and w_c, we can relax this condition (see Remark **??**).

Now, let

$$\eta(x;\xi) := \exp(\xi x)\,\mathfrak{w}(x).$$

Theorem 2 *Suppose that Assumption 3 holds, and let $\mathfrak{C}$ be a core given in (iii) of Assumption 3. Suppose also that for some interval Ξ on $\mathbb{R}$ which contains $\{0\}$, and for some continuous weighting function such that $q(x) > 0$ for any $x \in I$,*

$$E\left[\{(\mathcal{A}-\mathcal{A}_0)\,\eta(X_s;\xi)\}^2 \times q(X_s)\right] < \infty \quad \textit{for} \quad \forall \xi \in \Xi.$$

Then, there exists some test function $\varphi \in \mathfrak{C}$ satisfying

$$\Pr[\mathcal{A}\varphi(X_s) = \mathcal{A}_0\varphi(X_s)] < 1,$$

if and only if there exist some $\bar{\xi} > 0$ (which is arbitrarily close to zero) and some measurable subset $S(\bar{\xi})$ of $\{\xi : |\xi| \in [0, \bar{\xi}]\}$ with the Lebesgue measure of $S(\bar{\xi})$ non-zero, such that

$$\forall \xi \in S(\bar{\xi}),\ E\left[\{(\mathcal{A}-\mathcal{A}_0)\,\eta(X_s;\xi)\}^2 \times q(X_s)\right] > 0. \tag{27}$$

This theorem, together with Theorem 1, tells us that we only need to look at a family of parametrized functions, $\eta(x;\xi)$, in an arbitrarily small neighborhood of $\xi = 0$ for identifying the null and alternative hypotheses. The intuitive reason why this works is that any function in $\mathfrak{C}$ has a component *correlated* with a parametric family $\eta(x;\xi)$ in a certain sense. The following (rough) argument may clarify the meaning of this correlation. By the lemma above, we have $\varphi(x) \simeq \sum_{l=0}^{J-1} g_l(x)$ $(= \sum_{l=0}^{J-1} \gamma_l x^l \mathfrak{w}(x))$. If $\{\mathcal{A}-\mathcal{A}_0\}\varphi(x) \neq 0$, then it will also hold that $\{\mathcal{A}-\mathcal{A}_0\}\left(\sum_{l=0}^{J-1} g_l(x)\right) \neq 0$, which, by the linearity of the operators, implies that $\{\mathcal{A}-\mathcal{A}_0\}\,g_{l^*}(\cdot) \neq 0$ for some l^*. Now, since $\eta(x;\xi) = \sum_{l=0}^{\infty} \xi^l x^l \mathfrak{w}(x)/l!$,

$$\left(d^{l^*}/d\xi^{l^*}\right)(\mathcal{A}-\mathcal{A}_0)\,\eta(\cdot;\xi) \simeq (\mathcal{A}-\mathcal{A}_0)\,g_{l^*}(\cdot)$$

around $\xi = 0$ (take the interchangeability of $d^{l^*}/d\xi^{l^*}$ and $(\mathcal{A}-\mathcal{A}_0)$ as given for now). Therefore, if $\{\mathcal{A}-\mathcal{A}_0\}\varphi(\cdot) \neq 0$, then $(\mathcal{A}-\mathcal{A}_0)\,\eta(\cdot;\xi) \neq 0$ around $\xi = 0$.

Now, $(\mathcal{A}-\mathcal{A}_0)\,\eta(X_s;\xi) = 0$ for any ξ under the null hypothesis, and thus ξ is called a nuisance parameter. A similar technique to use nuisance parameters can be found in the so-called nuisance parameter approach (or Bierens' approach), where testing procedures for specification of parametric conditional moment functions (or regression functions) are developed. See Bierens (1982, 1990), De Jong (1996), Bierens and Ploberger (1997), Stinchcombe and White (1998), and Boning and Sowell (1999) (see also Andrews and Ploberger (1994), and Hansen (1996)). While the result demonstrated in Theorem 2 is (at least apparently) similar to results in those papers above, it is not an obvious extension, since we work with more complicated functional operators (differential operators define via conditional moment functions) instead of conditional moment functions.

The next theorem suggests another way to identify the hypotheses, whose difference from the previous theorem is twofold: the theorem below uses two nuisance parameters ξ_1 and ξ_2, and the moment (28) is not of the L^2-type as (27) is. While an additional nuisance parameter requires more computational effort in implementing the testing procedure, we will see that a test statistic based on the following theorem needs less conditions in deriving its asymptotic distribution than the one based on Theorem 2 (the mixing condition is no longer required).

Theorem 3 *Suppose that Assumption 3 holds, and let $\mathfrak{C}$ be a core given in (iii) of Assumption 3. Suppose also that for some intervals Ξ_1 and Ξ_2 on $\mathbb{R}$, each of which contains $\{0\}$, for some continuous weighting function such that $q(x) > 0$ for any $x \in I$, and for some $c \in (0, 1/2)$,*

$$E\left[(\mathcal{A} - \mathcal{A}_0)\,\eta(x;\xi_1) \times \exp(\xi_2 X_s)\, w_c(X_s)\, q(X_s)\right] \in (-\infty, \infty) \quad \textit{for} \ \ \forall (\xi_1, \xi_2) \in \Xi_1 \times \Xi_2.$$

Then, there exists some test function $\varphi \in \mathfrak{C}$ satisfying

$$\Pr\left[\mathcal{A}\varphi(X_s) = \mathcal{A}_0\varphi(X_s)\right] < 1,$$

if and only if there exist some $\left(\bar{\xi}_1, \bar{\xi}_2\right)$ with $\bar{\xi}_1 (> 0)$ and $\bar{\xi}_2 (> 0)$ (both of which are arbitrarily close to zero) and some measurable subset $S\left(\bar{\xi}_1, \bar{\xi}_2\right)$ of $\left\{(\xi_1, \xi_2) : |\xi_1| \in \left[0, \bar{\xi}_1\right],\ |\xi_2| \in \left[0, \bar{\xi}_2\right]\right\}$ with the Lebesgue measure of $S\left(\bar{\xi}_1, \bar{\xi}_2\right)$ non-zero, such that

$$\forall (\xi_1, \xi_2) \in S\left(\bar{\xi}_1, \bar{\xi}_2\right),\ E\left[(\mathcal{A} - \mathcal{A}_0)\,\eta(x;\xi_1) \times \eta(X_s;\xi_2)\, q(X_s)\right] \neq 0. \qquad (28)$$

The test function which appears in Theorems 2 and 3, $\eta(x;\xi) = \exp\{\xi x\}\,\mathfrak{w}(x)$, consists of two parts: the rescaled exponential function $\exp\{\xi x\}$ and the weighting function $\mathfrak{w}(x) = \exp\{-x^2/\}/\sqrt{2\pi}$. As is the case in the nuisance parameter approach, the exponential function can be replaced with some suitable analytic function such as $\cos(x) + \sin(x)$, $1/[1 + \exp\{c - x\}]$ (see Stinchcombe and While (1998)). $\mathfrak{w}(x)$ can also be replaced with some other *Freud-type* functions (see Balázs (2004)) (the same argument is applied to $\eta(X_s;\xi_2)$ in Theorem 3).

5 Test Statistics and Asymptotic Distributions

Theorems 1 and 2 suggest testing H_0 against H_1 by using an empirical counterpart of the following object :

$$M = \int_{\Xi} E\left[\left\{(\mathcal{A} - \mathcal{A}_0)\,\eta\,(X_s;\xi)\right\}^2 \times q\,(X_s)\right] d\xi.$$

Note that $M = 0$ if and only if H_0 is true. To construct the empirical counterpart of M and implement the testing procedure, we need to specify Ξ and q. As is seen in Theorem 2, any compact interval which contains $\{0\}$ can be used as Ξ. For simplicity, we set $\Xi = [-1, 1]$. q can also be chosen arbitrarily as long as the stated conditions in Theorem 2 are satisfied. We set q as the invariant density π of the process $\{X_s\}$. We construct a test statistic by replacing each component in M with its sample analog. Replace $\mathcal{A}\eta\,(\cdot;\xi)$, $\mathcal{A}_0\eta\,(\cdot;\xi)$ and $E\,[\cdot]$ with (13), (17) and the sample average (set $\varphi\,(\cdot) = \eta\,(\cdot;\xi)$), respectively. A natural choice of the estimator of $\pi\,(x)$ is $\hat{\pi}\,(x)$. This type of weighting is used in Fan and Li (1996), which obviously weights heavily towards the empirical mode of $\{X_s\}$, and is useful to overcome the random denominator problems in kernel estimation since the denominators of the kernel estimators for the generators are cancelled out (see the forms of the estimators (13) and (17)). Under these settings, $\hat{M}$ is written as

$$\hat{M} = \int_{\Xi} \hat{N}\,(\xi)\,d\xi, \quad \text{where}$$

$$\hat{N}\,(\xi) := \frac{1}{n}\sum_{i=1}^{n} \hat{\lambda}^2\,(X_{i\Delta};\xi)\,,$$

$$\hat{\lambda}\,(x;\xi) = \frac{1}{Th}\sum_{j=1}^{n-1} K\left(\frac{X_{j\Delta} - x}{h}\right)\Big\{\eta\left(X_{(j+1)\Delta};\xi\right) - \eta\,(X_{j\Delta};\xi) - \eta'\,(X_{j\Delta};\xi)\left[X_{(j+1)\Delta} - X_{j\Delta}\right] - \frac{1}{2}\eta''\,(X_{j\Delta};\xi)\left[X_{(j+1)\Delta} - X_{j\Delta}\right]^2\Big\}.$$

To see what $\hat{M}$ is trying to estimate, notice that $\hat{\lambda}$ is the weighted average of the residuals of the second order Taylor expansion of η. Ito's formula corresponding to diffusion processes claims that the third (or higher) order residuals should be negligible, and thus $\hat{\lambda}$ is expected to be close to zero under H_0. Therefore, our test can be interpreted as a test to check the validity of Ito's formula for a certain process.

To derive the asymptotic distribution of $\hat{M}$, we impose the following conditions:

Assumption 4 *(i) $\{X_s\}$ is stationary and ergodic with an invariant probability density π whose support is I.*

(ii) $\mathbf{M}_i(\cdot)$ *is at least twice continuously differentiable;* $|\mathbf{M}_i^{(k)}(\cdot)|$ *is uniformly bounded by some polynomial function for each* $k = 0, 1, 2$ *(for* $i = 1, 2$*);*

(iii) $\pi(\cdot)$ *is continuously differentiable, and* $\left|\pi^{(k)}(\cdot)\right|$ *is bounded by some polynomial function for each* $k = 0, 1$;

(iv) $\{X_s\}$ *is strong (α-) mixing with mixing coefficients* $\alpha(s)$ *which satisfy*

$$\alpha(s) \leq A \exp\{-\beta s\} \quad \textit{where} \quad a, \beta \in (0, \infty).$$

Theorem 4 *Suppose that the conditions in Theorems 1 and 2 hold; and Assumptions 2 and 4 also hold. Let* Z_0 *be a mean-zero Gaussian process in* $\mathbf{C}(\Xi)$ *with covariance kernel*

$$\Lambda_N(\xi, \xi') = \frac{1}{18} \int L^2(w)\, dw \int \left[\eta'''(y; \xi)\, \eta'''(y; \xi')\, \sigma^6(y)\, \pi^2(y)\right]^2 dy$$

where $L(w) := \int K(u-w) K(u)\, du$*. Under the null hypothesis* H_0*, as* $n, T \to \infty$ *and* $\Delta, h \to 0$ *with* $Th^2 \to \infty$*,* $n\Delta^2 \to 0$ *and* $\sqrt{\Delta \log(1/\Delta)}/h \to 0$

(i)

$$\hat{Z}(\xi) = \frac{\hat{N}(\xi) - (T/n^2 h)\, E_N(\xi)}{\sqrt{T^2/n^4 h}}$$

converges weakly to $Z_0(\xi)$ *in* $\mathbf{C}(\Xi)$ *where*

$$E_N(\xi) := \frac{1}{6} \int K^2(q)\, dp \times \int \left[\eta'''(x; \xi)\, \sigma^3(x)\right]^2 \pi^2(x)\, dx;$$

(ii)

$$\frac{n^2 h^{1/2}}{T} \hat{M} - \frac{1}{h^{1/2}} \int_{\Xi} E_N(\xi)\, d\xi \Longrightarrow m_0 \left(:= \int_{\Xi} Z_0(\xi)\, d\xi \right). \tag{29}$$

Implementation of the testing procedure To implement the testing procedure, we need to know the bias correction component $\int_{\Xi} E_N(\xi)\, d\xi$ in (29) as well as the distribution of m_0. While the former can be easily estimated, the asymptotic null distribution depends on a large number of unknown objects and is case-dependent. Thus, tabulation of critical values is infeasible. To proceed, we propose a simulation-based technique to approximate the distribution.

The bias component $E_N(\xi)$ can be estimated by

$$\hat{E}_N(\xi) = \frac{\int K^2(q)\, dp}{6} \times \frac{1}{n} \sum_{k=1}^{n} \left(\frac{\eta'''(X_{i\Delta}; \xi)}{\hat{\pi}(X_{i\Delta})} \right)^2 \left(\widehat{\sigma^2 \cdot \pi}(X_{i\Delta}) \right)^3$$

where

$$\widehat{\sigma^2 \cdot \pi}(x) = \frac{1}{n-1} \sum_{k=1}^{n-1} \frac{1}{h} K\left(\frac{X_{k\Delta} - x}{h} \right) \left[X_{(k+1)\Delta} - X_{k\Delta} \right]^2.$$

It can be easily shown that $\hat{E}_N(\xi) = E_N(\xi) + o_p\left(h^{1/2}\right)$ uniformly over ξ (under the same conditions as Theorem 2), since $\eta'''(x;\xi)$ is differentiable with respect to ξ and its derivative is uniformly bounded over x and ξ. Thus, we also have

$$\int_{\Xi} E_N(\xi)\,d\xi = \int_{\Xi} \hat{E}_N(\xi)\,d\xi + o_p\left(h^{1/2}\right). \tag{30}$$

Now, we outline how to approximate the distribution of m_0. The idea is to use another stochastic process $\tilde{Z}$ such that (i) $\tilde{Z}$ has the same limit as $\hat{Z}$; and (ii) it is computationally tractable and easy to simulate. Letting $\{u_j\}_{j=1}^n$ be an independent sequence (conditional on $\{X_{j\Delta}\}_{j=1}^n$) with $u_j = \sqrt{V_j} e_j$, $V(X_{j\Delta}) = \left(\int L^2(w)\,dw/18\right) \times \sigma^{12}(X_{j\Delta})\,\pi^3(X_{j\Delta})$, and $e_j \sim$ i.i.d. $N(0,1)$, define the stochastic process (indexed by ξ) $\tilde{Z}$ as

$$\tilde{Z}(\xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\eta'''(X_{j\Delta};\xi)\right]^2 u_j.$$

By the same argument as in the proof of Theorem 4, $\tilde{Z}$ also converges weakly to Z_0 in $\mathbf{C}(\Xi)$. The distribution of $\tilde{Z}$ can be simulated by the following steps:

1. Generate $\{e_j\}_{j=1}^n$ i.i.d. from $N(0,1)$ (using a random number generator), and set $\hat{u}_j = \sqrt{\hat{V}(X_{j\Delta})} e_j$ where $\hat{V}(X_{j\Delta})$ is an estimate of $V(X_{j\Delta})$;

2. Set $\hat{\tilde{Z}}(\xi) = (1/\sqrt{n}) \sum_{j=1}^n \eta'''(X_{j\Delta};\xi)\,\hat{u}_j$ for each $\xi \in \Xi$;

3. Compute
$$\hat{\tilde{m}} = \int_{\Xi} \hat{\tilde{Z}}(\xi)\,d\xi; \tag{31}$$

4. Repeat step 1-3 B times, say $B = 1000$.

This procedure gives a random sample $\{\hat{\tilde{m}}_b\}_{b=1}^B$ of B observations. A natural estimator for $V(x)$ in the first step is

$$\hat{V}(x) = \frac{\int L^2(w)\,dw}{18\hat{\pi}^3(x)} \left\{ \frac{1}{T} \sum_{k=1}^{n-1} \frac{1}{h} K\left(\frac{X_{k\Delta} - x}{h}\right) \left[X_{(k+1)\Delta} - X_{k\Delta}\right]^2 \right\}^6.$$

The uniform consistency of $\hat{V}(x)$ can be shown under suitable conditions (see Kanaya (2007)). Calculation of (30) and (31) might be excessively costly, and thus it may be reasonable to replace the integration by a discrete approximation:

$$\int_{[-1,1]} \hat{E}_N(\xi)\,d\xi \simeq \frac{1}{R} \sum_{r=1}^{R} \hat{E}_N(\xi_r)$$

where $\{\xi_r\}$ is an i.i.d.random sequence from an interval Ξ (from $U[-1,1]$ if $\Xi = [-1,1]$) or an appropriate pseudo random sequence. To test the null hypothesis at the level α, we would use the critical region:

$$\text{reject } H_0 \text{ when } \hat{M} \geq \hat{c}(\alpha) := \frac{T}{n^2 h} \int_{\Xi} \hat{E}_N(\xi)\, d\xi + \frac{T}{n^2 h^{1/2}} \widehat{\hat{m}}(1-\alpha; B)$$

where $\widehat{\hat{m}}(1-\alpha; B)$ is the $(1-\alpha)$-quantile of $\{\widehat{\hat{m}}_b\}_{b=1}^{B}$. The validity of this testing procedure can be verified by the same arguments as in Hansen (1996) and De Jong (1996).

We construct another test statistic based on the identification theorems 1 and 3. A sample analogue of

$$M_2 := \int_{\Xi_1} \int_{\Xi_2} \left\{ E\left[(\mathcal{A} - \mathcal{A}_0)\, \eta(X_s; \xi_1) \times \eta(X_s; \xi_2)\, q(X_s)\right] \right\}^2 d\xi_1 d\xi_2$$

should also work as a test statistic. By the same arguments as above, set q as π. Then, M_2 is consistently estimated by

$$\hat{M}_2 = \int_{\Xi_1} \int_{\Xi_2} \hat{N}_2^2(\xi_1, \xi_2)\, d\xi_1 d\xi_2 \quad \text{where}$$

$$\hat{N}_2(\xi_1, \xi_2) := \frac{1}{n} \sum_{i=1}^{n} \hat{\lambda}(x; \xi_1)\, \eta(X_{i\Delta}; \xi_2);$$

and $\hat{\lambda}$ is defined above. We also have a similar theorem, but only under weaker conditions (the mixing condition is not required).

Theorem 5 *Suppose that the same conditions as in Theorems 1 and 3 hold; and Assumption 2 and the conditions (i)-(iii) of Assumption 4 also hold. Then, under the null hypothesis* H_0*, as* $n, T \to \infty$ *and* $\Delta, h \to 0$ *with* $Th^2 \to \infty$, $n\Delta^2 \to 0$ *and* $\sqrt{\Delta \log(1/\Delta)}/h \to 0$

(i)

$$\hat{Z}_2(\xi_1, \xi_2) = \left(n/T^{1/2}\right) \hat{N}(\xi_1, \xi_2)$$

converges weakly to $Z_{2,0}(\xi_1, \xi_2)$*, where* Z_0 *is a mean-zero Gaussian process in* $\mathbf{C}(\Xi_1 \times \Xi_2)$ *with covariance kernel*

$$\Lambda_{N_2}\left((\xi_1, \xi_2), (\xi_1', \xi_2')\right) = \frac{1}{6} \int \eta(x; \xi_2)\, \eta(x; \xi_2')\, \eta'''(x; \xi_1)\, \eta'''(x; \xi_1')\, \sigma^6(x)\, \pi^5(x)\, dx;$$

(ii)

$$\left(n/T^{1/2}\right) \hat{M}_2 \Longrightarrow m_{2,0} \left(:= \int_{\Xi_1} \int_{\Xi_2} Z_{2,0}(\xi_1, \xi_2)\, d\xi_1 d\xi_2 \right).$$

Again, it is impossible to tabulate critical values since the asymptotic distribution is case-dependent. However, we can approximate the asymptotic critical values in the same manner as above, defining the stochastic process $\tilde{Z}_2$ as

$$\tilde{Z}_2(\xi_1, \xi_2) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta(X_{j\Delta}; \xi_2) \eta'''(X_{j\Delta}; \xi_1) u_{2,j}$$

where $\{u_{2,j}\}_{j=1}^{n}$ is an independent sequence (conditional on $\{X_{j\Delta}\}_{j=1}^{n}$) with $u_{2,j} = \sqrt{V_2(X_{j\Delta})} e_j$, $V_2(X_{j\Delta}) = \sigma^6(X_{j\Delta}) \pi^4(X_{j\Delta}) / 6$, and $e_j \sim$ i.i.d. $N(0,1)$.

6 Monte Carlo Results

In this section, we examine the finite-sample size and power properties of the proposed test. First, to see the size performance, we simulate the so-called CIR model suggested by Cox, Ross and Ingersoll (1985):

$$dX_s = \kappa(\mu - X_s) ds + \sigma \sqrt{X_s} dW_s$$

with $\kappa = 0.6$, $\mu = 0.09$ and $\sigma = 0.06$ (this parameter setting is used in Phillips and Yu (2005)). This specification of the drift and diffusion functions ensures that the data generating process is stationary and ergodic (all the other required conditions are also satisfied). The advantage of this model is that the transition density is known so that we can easily generate sample paths.[23] We set the time distance $\Delta = 1/250$, and sample sizes $n = 2500$, 5000, roughly corresponding to 10 and 20 years of data. The experiment is replicated 400 times. For each experiment, we compute the test statistic $\hat{M}$ and the critical values at the 5% and 10% significant levels using the simulation-based technique (the step 1-3 in the previous section is repeated 499 times). The integration with respect to ξ is computed by Monte Carlo integration using pseudo random numbers with $\Xi = [-1, 1]$ (the first 20 numbers of the Halton sequence are used). Throughout this experiment, we use the standard normal kernel. The bandwidth parameter h is chosen according to $h = c\hat{s}n^{-1/5}$ where $\hat{s}$ is the standard deviation of the observations. This choice of h satisfies the conditions in Theorem 4. To check the sensitivity of the proposed test with respect to the choice of bandwidth, we change h by setting different values of $c : c = 1, 2, 4$. $h = \hat{s}n^{-1/5}$ is optimal for i.i.d.data and used in Chapman and Pearson (2000). $h = 4\hat{s}n^{-1/5}$ is used in Stanton (1997). The results reported in Table 1 suggest that

[23] See, e.g., Cox, Ross and Ingersoll (1985) or Chapman and Pearson (2000). We use the `ncx2rnd` function of Matlab 7.0.

the bandwidth choice of $h = 4\hat{s}n^{-1/5}$ leads to better performance. The size performance of the proposed test seems satisfactory under this choice.

Next, we simulate the following (stochastic volatility) jump-diffusion model to examine the power property in the finite sample:

$$\begin{cases} dX_s = \sigma_s dW_s + \kappa_s dq_s \\ d\sigma_s^2 = \eta\left(\theta - \sigma_s^2\right) ds + v\sigma_s dZ_s \end{cases}$$

where q_s is a Poisson process with intensity λ; $\kappa_s \sim N\left(0, \vartheta^2\right)$, and the Brownian motions $\{W_s\}$ and $\{Z_s\}$ are uncorrelated. We set $(\theta, \eta, v) = (1, 0.01, 0.1)$ and $(\vartheta^2, \lambda) = (0.5, 0.2)$. This model and parameter settings, called "Moderate jumps," are used in Andersen, Bollerslev and Dobrev (2007). We generate sample paths of this model by the Euler scheme with a step length of $\delta = \Delta/10$ (paths of the latent volatility process are generated in the same way as above). Letting all the other settings be the same as above, we compute the percentage of rejection. According to the results in Table 2, the bandwidth $h = 4\hat{s}n^{-1/5}$ performs better again. The results also suggest that many observations are required to obtain reasonable power.

Table 1: percentage of rejection of the **true** H_0

Bandwidth $h =$	$\hat{s}n^{-1/5}$		$2\hat{s}n^{-1/5}$		$4\hat{s}n^{-1/5}$	
Nominal Size	5%	10%	5%	10%	5%	10%
$n = 2500$, $T = 10$	0.0250	0.0450	0.0325	0.0450	0.0300	0.0750
$n = 5000$, $T = 10$	0.0600	0.0675	0.0375	0.0900	0.0550	0.0975

Table 2: percentage of rejection of the **false** H_0

Bandwidth $h =$	$\hat{s}n^{-1/5}$		$2\hat{s}n^{-1/5}$		$4\hat{s}n^{-1/5}$	
Nominal Size	5%	10%	5%	10%	5%	10%
$n = 2500$, $T = 10$	0.0825	0.1750	0.1125	0.1925	0.1625	0.2100
$n = 5000$, $T = 10$	0.2050	0.3800	0.3050	0.4700	0.375	0.4825

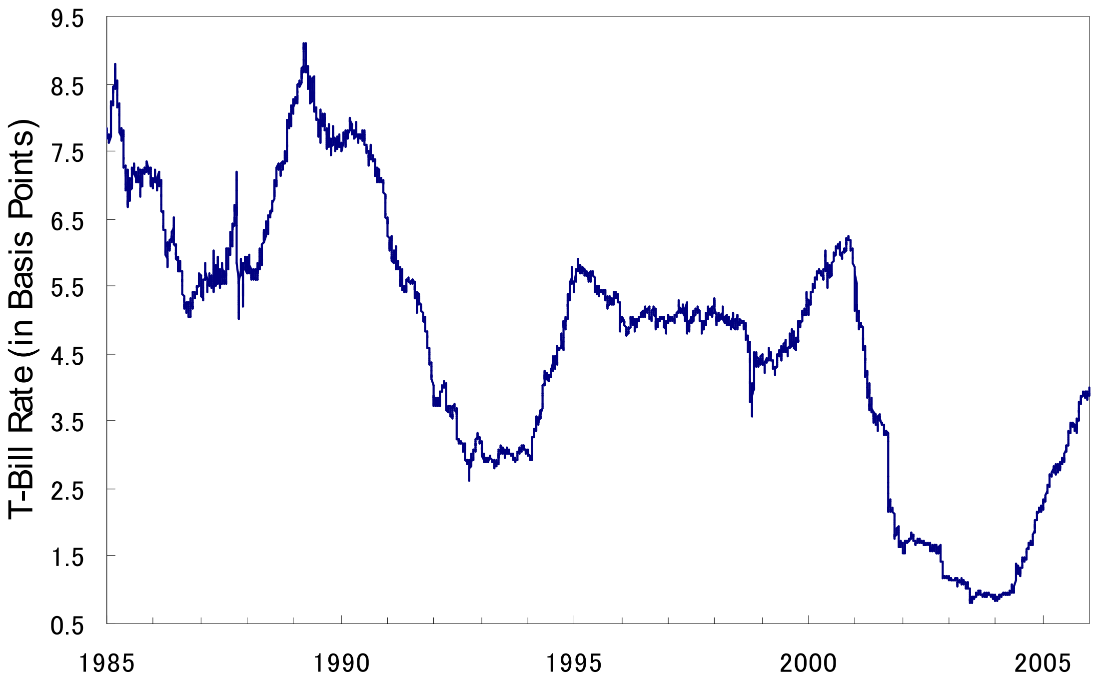

Figure 2: Three month Treasury Bill rate (the secondary market) daily observations from 01/02/1985 to 12/30/2005 (annualized yield).

7 Empirical Applications

I apply the proposed method to short term interest rates and foreign exchange rates to examine the validity of the diffusion hypothesis. As a proxy for the short term rates, I use daily values of the secondary market yields on three month Treasury Bills from 01/02/1985 to 12/30/2005 (the number of observations is $n = 5250$).[24] This three month Treasury Bill rate is also used in Stanton (1997) with a different sample period. Figure 2 shows a time-series plot of the short term rates. Figure 3 shows a time-series plot of daily observations of the Norwegian Kroner-British Pound exchange rate from 01/02/1985 to 12/30/2005 (the sample size is $n = 5309$). This Norwegian exchange rate data was chosen because it is said to display the generic characteristics of foreign exchange data (Craine, Lochstoer and Syrveit (2000)).

Tests are conducted with the bandwidth $h = 4\hat{s}n^{-1/5}$. We measure time in years and set the time distance $\Delta = 1/252$, so that no specific adjustment is made for weekends or holidays (all the other settings are the same as in the previous section). Test results are reported in Table 3.

[24] The obtained values are converted from discounts to annualized interest rates.

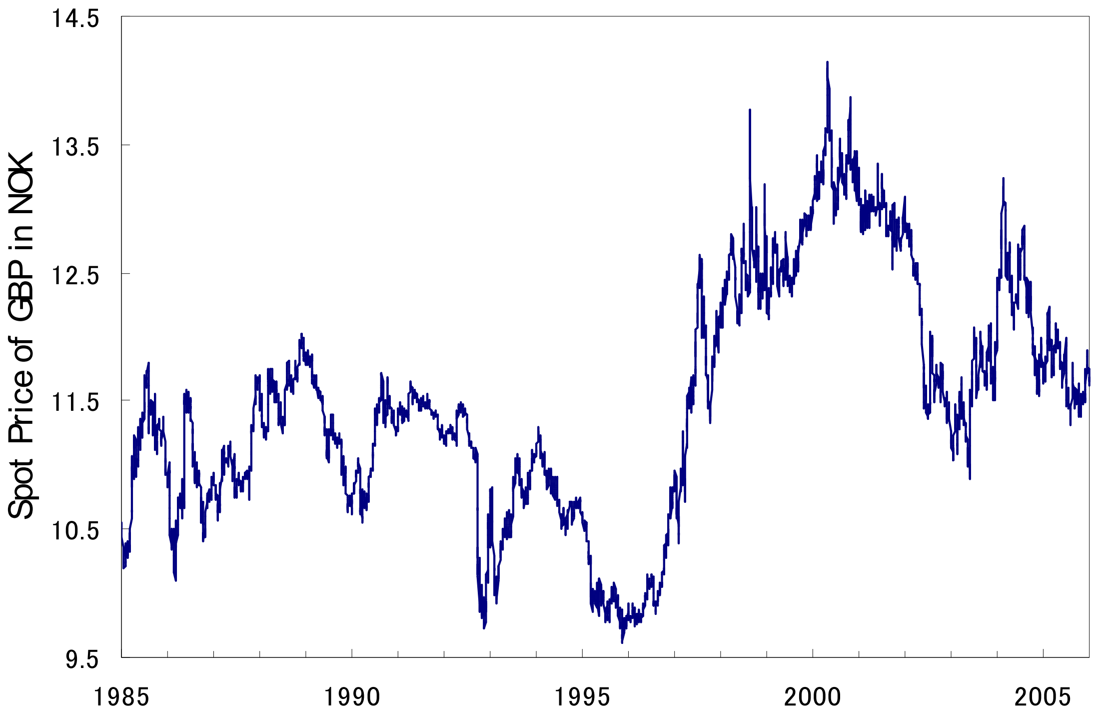

Figure 3: The Norweigian Kroner - British Poound exchange rate from 01/02/1985 to 12/30/2005.

We find that (i) the diffusion hypothesis seems appropriate for this particular data of the short term rates; but (ii) it may be inappropriate for the exchange rate data. The first finding seems consistent with previous studies: e.g., Aït-Sahalia (1996) does not reject a flexible parametric diffusion model in his specification test. The rejection of the diffusion hypothesis for the exchange rate data may be caused by its complicated dynamic behaviors.

Table 3: tests for the diffusion hypothesis

	3 month T-Bill Rate	NOK/GBP Exchange Rate
Test statistic $\hat{M}$	5.415E-10	8.739E-8
p-value	0.5300	0.0000

8 Conclusion

We have proposed a new statistical testing procedure to examine the diffusion hypothesis. While the infinitesimal generator has been known to be a useful and convenient tool to characterize the Markov process, it has not been used extensively in the econometric literature.

Our testing procedure is based on two novel results: (i) a new theorem to identify diffusion processes through infinitesimal generators; (ii) asymptotic results for the proposed test statistics. We also found that our simulation studies support the theoretical results.

Various related issues and extensions are of interest for future research. As argued earlier, our framework could be used to construct various specification tests of parametric/semiparametric Markov processes. For example, a specification test of the parametric volatility function $\sigma(\cdot;\theta)$ of a jump-diffusion model:

$$dX_s = \mu\left(X_{s_-}\right) ds + \sigma\left(X_{s_-};\theta\right) dW_s + dJ_s.$$

The validity of the parametric restriction of $\sigma(\cdot;\theta)$ can be examined with the drift function $\mu(\cdot)$ and the jump component J_s unspecified. In fact, any combination of parametric/nonparametric restrictions for the three components is possible, since the infinitesimal generator can be easily estimated in all cases. Another interesting extension would be to consider weakly non-stationary processes (as in Bandi and Phillips (2003, 2007) and Karlsen and Tjøstheim (2001)). Our identification scheme through infinitesimal generators works without stationarity, and thus problems can all be traced to deriving relevant asymptotic results, where the local-time approach (developed by Bandi and Phillips (2003, 2007), Bandi and Nguyen (2003), Park (2007) and others) might be useful.

References

[1] Aït-Sahalia, Y. (1996) "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 64, 527-560.

[2] Aït-Sahalia, Y. (1999) "Transition Densities for Interest Rate and Other Nonlinear Diffusions," *Journal of Finance*, 54, 1361-1395.

[3] Aït-Sahalia, Y. (2002a) "Telling from Discrete Data Whether the Underlying Continuous-Time Model Is a Diffusion," *Journal of Finance*, 57, 2075-2112.

[4] Aït-Sahalia, Y. (2002b) "Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach," *Econometrica*, 70, 223-262.

[5] Aït-Sahalia, Y., Bickel, P.J. and Stocker, T.M. (2001) "Goodness-of-fit tests for kernel regression with an application to option implied volatilities," *Journal of Econometrics*, 105, 363-412.

[6] Aït-Sahalia, Y., Fan, J. and Peng, H. (2006) "Nonparametric Transition-Based Tests for Jump-Diffusions," Working paper, Princeton University.

[7] Aït-Sahalia, Y. and Jacod, J. (2007) "Testing for Jumps in a Discretely Observed Process," forthcoming in *Annals of Statistics*.

[8] Aït-Sahalia, Y. and Mykland, P.A. (2007) "An Analysis of Hansen-Scheinkman Moment Estimators for Discretely and Randomly Sampled Diffusions," forthcoming in *Journal of Econometrics*.

[9] Andersen, T.G., Bollerslev, T. and Dobrev, D. (2007) "No-arbitrage Semi-Martingale Restrictions for Continuous-Time Volatility Models Subject to Leverage Effects, Jumps and I.I.D. Noise: Theory and Testable Distributional Implications," *Journal of Econometrics*, 138, 125-180.

[10] Andrews, D.W.K and Ploberger, W. (1994) "Optimal Test When a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, 1383-1414.

[11] Applebaum, D. (2004) *Lévy Processes and Stochastic Calculus*, Cambridege University Press.

[12] Back, K. (1991) "Asset Pricing for General Processes," *Journal of Mathematical Economics*, 20, 371-395.

[13] Balázs, K. (2004) "Weighted Simultaneous Approximation by Lagrange Interpolation on the Real Line," *Computers and Mathematics with Applications*, 45, 1345-1349.

[14] Bandi, F.M. and Nguyen, T.H. (2003) "On the functional Estimation of Jump-Diffusion Models," *Journal of Econometrics*, 116, 293-328.

[15] Bandi, F. M. and Phillips, P.C.B. (2003) "Fully Nonparametric Estimation of Scholar Diffusion Models," *Econometrica*, 71, 241-283.

[16] Bandi, F. M. and Phillips, P.C.B. (2007) "A Simple Approach to the Parametric Estimation of Potentially Nonstationary Diffusions," *Journal of Econometrics*, 137, 354-395.

[17] Barndorff-Nielsen, O. E. and Shephard, N. (2001) "Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of their Uses in Financial Economics," *Journal of Royal Statistics Society B*, 63, 167-241.

[18] Barndorff-Nielsen, O. E. and Shephard, N. (2006) "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4, 1-30.

[19] Bertoin, J. (1998) *Lévy Processes*, Cambridge University Press.

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

[21] Boning, Wm. B. and Sowell F. (1999) "Optimality for the Integrated Conditional Moment Test," *Econometric Theory*, 15, 710-718.

[22] Brandt, M. and P. Santa-Clara (2002) "Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets," *Journal of Financial Economics*, 63, 161-210.

[23] Bierens, H.J. (1982) "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105-134.

[24] Bierens, H.J. (1990) "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1485.

[25] Bierens, H.J. and Ploberger, W. (1997) "Asymptotic Theory of Integrated Conditional Moments Tests," *Econometrica*, 65, 1129-1151.

[26] Black, Fischer and Myron S. Scholes (1973) "The Pricing of Option and Corporative Liabilities," *Journal of Political Economy*, 81, 637-654.

[27] Carr, P., German, H., Madan, D.B. and Yor, M. (2002) "The Fine Structure of Asset Returns: An Empirical Investigation," *Journal of Business*, 75, 305-332.

[28] Chapman, D.A. and Pearson, N.D. (2000) "Is the Short rate Drift Actually Nonlinear?" *Journal of Finance*, 55, 355-388.

[29] Conley, T.G., Hansen, L.P., Luttmer, E.G.J. and Scheinkman, J.A. (1997) "Short-Term Interest Rates as Subordinated Diffusions," *The Review of Financial Studies*, 10, 525-577.

[30] Constantinides, G.M. (1990) "Habit Formation: A Resolution of the Equity Premium Puzzle," *The Journal of Political Economy*, 98, 519-543.

[31] Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985) "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.

[32] Cox, J.C. and Ross, S.A. (1976) "The valuation of options for alternative stochastic processes," *Journal of Financial Economics*, 1976, 3, 145-166.

[33] Craine, R., Lochstoer, L. and Syrveit, K. (2000) "Estimation of a Stochastic-Volatility Jump-Diffusion Model," *Revista de Analisis Economico*, 2000, 61-87.

[34] Darolles, S., Florens, J.P. and Gouriéroux, C. (2004) "Kernel-Based Nonlinear Canonical Analysis and Time Reversibility," *Journal of Econometrics* 119, 323-353.

[35] De Jong, R.M. (1996) "The Bierens Test under Data Dependence," *Journal of Econometrics*, 72, 1-32.

[36] Duffie, D. and Glynn, P. (2004) "Estimation of Continuous-Time Markov Processes Sampled at Random Time Interval," *Econometrica*, 72, 1773-1808.

[37] Dynkin, E.B. (1956) "Markov Processes and Semigroups of Operators," *Theory of Probability and its Applications*, 1, 149-155.

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

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

[40] Fan, J. (2006a) "A Selective Overview of Nonparametric Methods in Financial Econometrics," *Statistical Science*, 20, 317-337.

[41] Fan, J. (2006b) "Rejoinder: A Selective Overview of Nonparametric Methods in Financial Econometrics," *Statistical Science*, 20, 351-357.

[42] Fan, Y. and Li, Q. (1996) "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64, 865-890.

[43] Friedman, A (1975) *Stochastic Differential Equations and Applications*, Vol. 1, Academic Press.

[44] Froot, K.A. and Obstfeld, M. (1991) "Exchange-Rate Dynamics Under Stochastic Regime Shifts: A Unified Approach," *Journal of International Economics*, 31, 203-229.

[45] Florens, J. P., Renault, E. and N. Touzi (1998) "Testing for Embeddability by Stationary Reversible Continuous-Time Markov Processes," *Econometrics Theory* 14, 744-769.

[46] Gao, J. and King, M. (2004) "Adaptive Testing in Continuous-Time Diffusion Models," *Econometrics Theory*, 20, 844-882.

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

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

[49] Hansen, L.P. and Scheinkman, J.A. (1995) "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes," *Econometrica*, 63, 767-804.

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

[51] Heath, D., Jarrow, R. and Morton, A. (1992) "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation," *Econometrica*, 60, 77-105.

[52] Hong, Y. and Li, H. (2005) "Nonparametric Specification Testing for Continuous-Time Models with Application to Term Structure of Interest Rates," *The Review of Financial Studies*, 18, 37-84.

[53] Ikeda, N. and Watanabe, S. (1981) *Stochastic Differential Equations and Diffusion Processes*, North Holland.

[54] Jacob, N. (2001) *Pseudo Differential Operators and Markov Processes I*, Imperial College Press.

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

[56] Johannes, M. (2004) "The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models," *Journal of Finance*, 59, 227-260.

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

[58] Kanaya, S. (2007) "Uniform Convergence of Kernel-Based Nonparametric Estimators for Diffusion Processes: A Damping Function Approach," Working paper, University of Wisconsin-Madison

[59] Kanaya, S. and Kristensen, D. (2007) "Estimation of Stochastic Volatility Models by Nonparametric Filtering," Working paper, University of Wisconsin-Madison and Columbia University.

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

[61] Kristensen, D. (2006) "Nonparametric Filtering of the Realised Spot Volatility: A Kernel-based Approach," Working paper, Columbia University.

[62] Krugman, P.R. (1991) "Target Zones and Exchange Rate Dynamics," *The Quarterly Journal of Economics*, 106, 669-682.

[63] Kutoyants, Y.A. (2004) *Statistical Inference for Ergodic Diffusion Processes*, Springer-Verlag

[64] Li, F. (2007) "Testing the Parametric Specification of the Diffusion Function in a Diffusion Process," *Econometric Theory*, 23, 221-250.

[65] Lo, A.W. (1988) "Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data," *Econometric Theory*, 4, 231-247.

[66] Merton, R.C. (1973) "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867-888.

[67] Merton, R.C. (1974) "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance*, 29, 449-470.

[68] Merton, R.C. (1976) "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics*, 1976, 3, 125-144.

[69] Merton, R.C. (1980) "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 8, 323-361.

[70] Nicolau, J. (2002) "Stationary Processes that Look Like Random Walks – The Bounded Random Walk Process in Discrete and Continuous Time," *Econometric Theory*, 18, 99-118.

[71] Nicolau, J. (2005) "Processes with Volatility-Induced Stationarity: An Application for Interest Rates," *Statistica Neerlandica*, 59, 376-396.

[72] Park, J.Y. (2007) "The Spatial Analysis of Time Series," Working paper, Texas A & M University.

[73] Pedersen, A.R. (1995) "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equation Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22, 55-71.

[74] Peters, R.T., and de Vilder, R.G. (2006) "Testing the Continous Semimartingale Hypothesis for the S&P 500," *Journal of Business and Economic Statistics*, 24, 444-454.

[75] Phillips, P.C.B and Yu, J. (2005) "Jackknifing Bond Option Prices," *The Review of Financial Studies*, 18, 707-742.

[76] Phillips, P.C.B and Yu, J. (2006) "Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics," *Statistical Science*, 20, 338-343.

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

[78] Schaumburg, E. (2004) "Estimation of Markov Processes with Levy Type Generators," Working paper, Kellogg School of Management, Northwestern University.

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

[80] Stanton, R. (1997) "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, 52, 1973-2002.

[81] Stinchcombe M.B., and White H. (1998) "Consistent Specification Testing With Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295-325.

[82] Stramer, O. and Tweedie, R.L. (1999) "Langevin-Type Models I: Diffusions with Given Stationary Distributions and their Discretizations," *Methodology and Computing in Applied Probability*, 1, 286-306.

[83] Szabados, J. (1997) "Weighted Lagrange and Herimite-Fejér Interpolation on the Real Line," *Journal of Inequalities and Applications*, 1, 99-123.

[84] Todorov, V. (2006) "Econometric Analysis of Jump-Driven Stochastic Volatility Models," Working paper, Duke University.

[85] Vasicek, O. (1977) "An Equilibrium Characterisation of the Term Structure," *Journal of Financial Economics*, 5, 177-188.

[86] Xu, K. (2006) "Empirical Likelihood-Based Inference for Nonlinear Diffusions," Working paper, Yale University.

[87] Zhou, H. (2003) "Itô Conditional Moment Generator and the Estimatioin of Short-Rate Processes," *Journal of Financial Econometrics*, 1, 250-271.

A Proofs of theorems and lemmas for identifying the hypotheses

Proof of Theorem 1. The "if" part is obvious, and thus we prove the "only if" part.

Suppose that $\mathfrak{D}(\mathcal{A}) \cap \mathfrak{D}(\mathcal{A}_0) \neq \varnothing$ (otherwise, regard $\mathcal{A}$ ($\mathcal{A}_0$) as an extension of $\mathcal{A}_0$ ($\mathcal{A}$) and the proof is done). Suppose also that (23) does not hold: i.e., there exists some $\varphi \in \mathfrak{D}(\mathcal{A}) \cap \mathfrak{D}(\mathcal{A}_0)$ such that $||\mathcal{A}\varphi - \mathcal{A}_0\varphi|| > 0$, and thus the process is not a diffusion. By definition of the core, for each $\varphi \in \mathfrak{D}(\mathcal{A}) \cap \mathfrak{D}(\mathcal{A}_0)$, there exist a sequence $\{\varphi_l\}$ ($\varphi_l \in \mathfrak{C}$) such that

$$\lim_{l\to 0} ||\varphi_l - \varphi|| \to 0; \quad \lim_{l\to 0} ||\mathcal{A}\varphi_l - \mathcal{A}\varphi|| \to 0. \tag{32}$$

Let $\mathbf{w}$ be some weighting function such that $\mathbf{w}$ is continuous on $\mathbb{R}$ and $\mathbf{w}(x) > 0$ for any $x \in I$; and

$$\max\left\{\sup_{x\in\mathbb{R}} |\mathbf{M}_1(x)\mathbf{w}(x)|, \ \sup_{x\in\mathbb{R}} |\mathbf{M}_2(x)\mathbf{w}(x)|\right\} < C_{\mathbf{w}}. \tag{33}$$

Such $\mathbf{w}$ should exist by Assumption 1. Now, since the process is not a diffusion, the only requirement on $\mathfrak{D}(\mathcal{A}_0)$ is that it consists of (at least) twice differentiable functions. Note that $\mathfrak{C}$ is dense in the space of (at least) twice differentiable functions with respect to the norm $||f||_{\#} := \sup_{x\in\mathbb{R}} |f(x)| + \sup_{x\in\mathbb{R}} |f'(x)| + \sup_{x\in\mathbb{R}} |f''(x)|$, and thus for any $\varphi \in \mathfrak{D}(\mathcal{A}_0)$, there exists a sequence of functions $\{\tilde{\varphi}_l\}$ ($\tilde{\varphi}_l \in \mathfrak{C}$) $\lim_{l\to 0} ||\tilde{\varphi}_l - \varphi||_{\#} \to 0$. By (33), $\{\tilde{\varphi}_l\}$ satisfies

$$\lim_{l\to 0} ||\mathcal{A}_0\tilde{\varphi}_l - \mathcal{A}_0\varphi||_{\mathbf{w}} \to 0. \tag{34}$$

If $||\mathcal{A}\varphi - \mathcal{A}_0\varphi|| > 0$, we also have

$$\sup_{x\in\mathbb{R}} |\mathcal{A}\varphi(x) - \mathcal{A}_0\varphi(x)|\,\mathbf{w}(x) > 0,$$

since $\mathbf{w}(x) > 0$ for any x. In what follow we write $||f||_{\mathbf{w}} := \sup_{x\in\mathbb{R}} |f(x)|\,\mathbf{w}(x)$ (the weighted uniform norm). By triangle inequalities,

$$\begin{aligned} 0 &< ||\mathcal{A}\varphi - \mathcal{A}_0\varphi||_{\mathbf{w}} \\ &\leq ||\mathcal{A}\varphi - \mathcal{A}\varphi_l||_{\mathbf{w}} + ||\mathcal{A}\varphi_l - \mathcal{A}_0\varphi_l||_{\mathbf{w}} + ||\mathcal{A}_0\varphi_l - \mathcal{A}_0\tilde{\varphi}_l||_{\mathbf{w}} + ||\mathcal{A}_0\tilde{\varphi}_l - \mathcal{A}_0\varphi||_{\mathbf{w}} \quad (35) \\ &= e_l + ||\mathcal{A}\varphi_l - \mathcal{A}_0\tilde{\varphi}_l||_{\mathbf{w}} \quad (36) \end{aligned}$$

where $e_l := ||\mathcal{A}\varphi - \mathcal{A}\varphi_l||_{\mathbf{w}} + ||\mathcal{A}_0\tilde{\varphi}_l - \mathcal{A}_0\varphi||_{\mathbf{w}}$. Note that the last equality holds since

$$||\mathcal{A}\varphi_l - \mathcal{A}_0\varphi_l||_{\mathbf{w}} \leq ||\mathcal{A}\varphi_l - \mathcal{A}_0\varphi_l|| \times ||\mathbf{w}|| = 0 \ \text{ for } \varphi_l \in \mathfrak{C},$$

and that e_l converges to zero as $l \to \infty$ (by (32) and (34)).

We below show the convergence of the third term of (35). To this end, we consider two technical devices: Friedrichs's mollifier ρ (kernel function) and a (smooth) truncation function λ. Letting $\{b_m\}$ be a sequence of positive numbers such that $b_m \to 0$ as $m \to 0$, ρ and λ are defined as follows:

$$\rho(x) := \begin{cases} c_\rho \exp\left\{\dfrac{1}{x^2-1}\right\} & \text{if } |x| < 1; \\ 0 & \text{if } |x| \geq 1 \end{cases}$$

where c_ρ is determined to satisfy $\int_{-\infty}^{\infty} \rho(x)\, dx = 1$. Define λ by

$$\lambda(x; b_m) := \begin{cases} 1 & \text{if } |x| < 1/2b_m; \\ \exp\left\{-\exp\left\{\dfrac{-1}{(|x|-1/2b_m)^2 / (|x|-1/b_m)^2}\right\}\right\} & \text{if } |x| \in (1/2b_m, 1/b_m); \\ 0 & \text{if } |x| \geq 1/b_m. \end{cases}$$

Note that both $\rho(\cdot)$ and $\lambda(\cdot; b_m)$ is infinitely differentiable on $x \in \mathbb{R}$ (for each b_m). Given these functions, we construct four sequences of functions for each l : $\{\phi_{l,m} : m = 1, 2, \ldots\}$, $\{\eta_{l,m} : m = 1, 2, \ldots\}$, $\{\phi_{l,m} : m = 1, 2, \ldots\}$ and $\{\tilde{\eta}_{l,m} : m = 1, 2, \ldots\}$, defined as follows:

$$\begin{aligned} \phi_{l,m}(x) &:= \lambda(x; b_m)\, \varphi_l(x); \\ \eta_{l,m}(x) &:= \lambda(x; b_m) \int_{-\infty}^{\infty} \frac{1}{b_m} \rho\left(\frac{x-y}{b_m}\right) \varphi_l(y)\, dy; \\ \tilde{\eta}_{l,m}(x) &:= \lambda(x; b_m) \int_{-\infty}^{\infty} \frac{1}{b_m} \rho\left(\frac{x-y}{b_m}\right) \tilde{\varphi}_l(y)\, dy; \text{ and} \\ \tilde{\phi}_{l,m}(x) &:= \lambda(x; b_m)\, \tilde{\varphi}_l(x). \end{aligned}$$

where the first two functions are intended to approximate φ_l and the others $\tilde{\varphi}_l$. Again, by triangle inequalities,

$$\begin{aligned} & ||\mathcal{A}_0\varphi_l - \mathcal{A}_0\tilde{\varphi}_l||_{\mathbf{w}} \\ & \leq ||\mathcal{A}_0\varphi_l - \mathcal{A}_0\phi_{l,m}||_{\mathbf{w}} + ||\mathcal{A}_0\phi_{l,m} - \mathcal{A}_0\eta_{l,m}||_{\mathbf{w}} \\ & + ||\mathcal{A}_0\eta_{l,m} - \mathcal{A}_0\tilde{\eta}_{l,m}||_{\mathbf{w}} + \left|\left|\mathcal{A}_0\tilde{\eta}_{l,m} - \mathcal{A}_0\tilde{\phi}_{l,m}\right|\right|_{\mathbf{w}} + \left|\left|\mathcal{A}_0\tilde{\phi}_{l,m} - \mathcal{A}_0\tilde{\varphi}_l\right|\right|_{\mathbf{w}}. \end{aligned} \tag{37}$$

I show below that the right-hand side tends to zero if we let $l \to \infty$ and $m \to \infty$ (the order of the limits does not matter). The following two lemmas, whose proofs are given below, present the convergence results for the first and fifth terms of the right-hand side of (37), and the second and fourth terms, respectively:

Lemma 2 *As $l, m \to \infty$, (i) $||\mathcal{A}_0\varphi_l - \mathcal{A}_0\phi_{l,m}||_{\mathbf{w}} \to 0$; and (ii) $\left|\left|\mathcal{A}_0\tilde{\phi}_{l,m} - \mathcal{A}_0\tilde{\varphi}_l\right|\right|_{\mathbf{w}} \to 0$.*

Lemma 3 *As $l, m \to \infty$, (i) $||\mathcal{A}_0\phi_{l,m} - \mathcal{A}_0\eta_{l,m}||_{\mathbf{w}} \to 0$; and (ii) $\left|\left|\mathcal{A}_0\tilde{\eta}_{l,m} - \mathcal{A}_0\tilde{\phi}_{l,m}\right|\right|_{\mathbf{w}} \to 0$.*

It remains to show the convergence of the third term of the right-hand side of (37). Write

$$\eta_{\infty,m}(x) := \lambda(x; b_m) \int_{-\infty}^{\infty} \frac{1}{b_m}\rho\left(\frac{x-y}{b_m}\right)\varphi(y)\,dy; \text{ and}$$
$$\tilde{\eta}_{\infty,m}(x) := \lambda(x; b_m) \int_{-\infty}^{\infty} \frac{1}{b_m}\rho\left(\frac{x-y}{b_m}\right)\tilde{\varphi}(y)\,dy,$$

and look at the following inequality:

$$\begin{aligned} &||\mathcal{A}_0\eta_{l,m} - \mathcal{A}_0\tilde{\eta}_{l,m}||_{\mathbf{w}} \\ &\le ||\mathcal{A}_0\eta_{l,m} - \mathcal{A}_0\eta_{\infty,m}||_{\mathbf{w}} + ||\mathcal{A}_0\eta_{\infty,m} - \mathcal{A}_0\tilde{\eta}_{\infty,m}||_{\mathbf{w}} + ||\mathcal{A}_0\tilde{\eta}_{\infty,m} - \mathcal{A}_0\tilde{\eta}_{l,m}||_{\mathbf{w}}, \end{aligned}$$

where $||\mathcal{A}_0\eta_{\infty,m} - \mathcal{A}_0\tilde{\eta}_{\infty,m}||_{\mathbf{w}} = 0$ since $\eta_{\infty,m} = \tilde{\eta}_{\infty,m}$; the convergence results of $||\mathcal{A}_0\eta_{l,m} - \mathcal{A}_0\eta_{\infty,m}||_{\mathbf{w}}$ and $||\mathcal{A}_0\tilde{\eta}_{\infty,m} - \mathcal{A}_0\tilde{\eta}_{l,m}||_{\mathbf{w}}$ are given in the proof of Lemma 3. From these arguments, we have shown

$$||\mathcal{A}_0\varphi_l - \mathcal{A}_0\tilde{\varphi}_l||_{\mathbf{w}} \to 0 \quad \text{as} \quad l \to \infty. \tag{38}$$

By (36) and (38), we have

$$0 < ||\mathcal{A}\varphi - \mathcal{A}_0\varphi||_{\mathbf{w}} \le e_l + ||\mathcal{A}\varphi_l - \mathcal{A}_0\tilde{\varphi}_l||_{\mathbf{w}} \to 0,$$

a contradiction. ■

Proof of Lemma 2. (ii) can be shown in exactly the same manner as (i), and thus we only prove (i). By triangle inequalities,

$$\begin{aligned} &||\mathcal{A}_0\varphi_l - \mathcal{A}_0\phi_{l,m}||_{\mathbf{w}} \\ &\le \underbrace{\sup_{x \le 1/(2b_m)} |\mathcal{A}_0\varphi_l(x) - \mathcal{A}_0\phi_{l,m}(x)|\,\mathbf{w}(x)}_{=0} + \sup_{x > 1/(2b_n)} |\mathcal{A}_0\varphi_l(x) - \mathcal{A}_0\phi_{l,m}(x)|\,\mathbf{w}(x) \\ &\le R_1 + R_2 + R_3 \end{aligned}$$

where

$$\begin{aligned} &R_1 := \sup_{x > 1/(2b_m)} |\mathcal{A}_0\varphi_l(x) - \mathcal{A}_0\varphi(x)|\,\mathbf{w}(x); \quad R_2 := \sup_{x > 1/(2b_m)} |\mathcal{A}_0\varphi(x) - \mathcal{A}_0\phi_{\infty,m}(x)|\,\mathbf{w}(x); \\ &R_3 := \sup_{x > 1/(2b_m)} |\mathcal{A}_0\phi_{\infty,m}(x) - \mathcal{A}_0\phi_{l,m}(x)|\,\mathbf{w}(x); \quad \text{and} \quad \phi_{\infty,m}(x) := \lambda(x; b_m)\varphi(x). \end{aligned}$$

I show below that each term converges to zero.

First look at R_1 :

$$R_1 \leq \sup_{x>1/(2b_m)} |\mathbf{M}_1(x)\mathbf{w}(x)| \times \sup_{x>1/(2b_m)} |\varphi_l'(x) - \varphi'(x)| + \sup_{x>1/(2b_m)} |\mathbf{M}_2(x)\mathbf{w}(x)|/2 \times \sup_{x>1/(2b_m)} |\varphi_l''(x) - \varphi''(x)| \tag{39}$$

By the mean-value theorem, we have

$$\begin{cases} \varphi_l'(x) = \varphi_l(x+1-\epsilon_{x,l}) - \varphi_l(x-\epsilon_{x,l}) & \text{for some } |\epsilon_{x,l}| \leq 1; \\ \varphi'(x) = \varphi(x+1-\epsilon_x) - \varphi(x-\epsilon_x) & \text{for some } |\epsilon_x| \leq 1, \end{cases}$$

and thus

$$\sup_{|x|>1/(2b_m)} |\varphi_l'(x) - \varphi'(x)| \leq \underbrace{\sup_{|x|>1/(2b_m)} |\varphi_l(x-\epsilon_{x,l}) - \varphi(x-\epsilon_x)|}_{\equiv R_{11}} + \underbrace{\sup_{|x|>1/(2b_m)} |\varphi_l(x+1-\epsilon_{x,l}) - \varphi(x+1-\epsilon_x)|}_{\equiv R_{12}}$$

where R_{11} is bounded by

$$R_{11} \leq \sup_{|x|>1/(2b_m)} |\varphi_l(x-\epsilon_{x,l}) - \varphi(x-\epsilon_{x,l})| + \sup_{|x|>1/(2b_m)} |\varphi(x-\epsilon_{x,l}) - \varphi(x-\epsilon_x)|$$
$$\leq ||\varphi_l - \varphi|| + 2 \sup_{|x|>1/(2b_m)-1} |\varphi(x)| \to 0 \ \text{ as } l, m \to \infty.$$

The convergence of R_{12} follows from the same arguments as R_{11}, and thus,

$$\sup_{|x|>1/(2b_m)} |\varphi_l'(x) - \varphi'(x)| \to 0 \ \text{ as } \ l, m \to 0. \tag{40}$$

By the same arguments (but more tedious), it also holds that

$$\sup_{x>1/(2b_m)} |\varphi_l''(x) - \varphi''(x)| \to 0 \ \text{ as } \ l, m \to 0. \tag{41}$$

We have chosen $\mathbf{w}$ so that $\mathbf{M}_1(x)\mathbf{w}(x)$ and $\mathbf{M}_2(x)\mathbf{w}(x)$ are uniformly bounded, which together with (39), (40), and (41) implies that $F_1 \to 0$ as $l, m \to 0$.

The bound of R_2 is given by

$$R_2 \leq \sup_{x>1/(2b_m)} |\mathbf{M}_1(x)w(x)| \times \sup_{x>1/(2b_m)} \left|\frac{d}{dx}\varphi(x)(1-\lambda(x;b_m))\right| + \sup_{x>1/(2b_m)} |\mathbf{M}_2(x)w(x)|/2 \times \sup_{x>1/(2b_m)} \left|\frac{d^2}{dx^2}\varphi(x)(1-\lambda(x;b_m))\right|.$$

To show the convergence of the right-hand side, notice that φ' and φ'' are expressed as

$$\varphi'(x) = \varphi(x+1-\epsilon_x) - \varphi(x-\epsilon_x) \quad \text{for some } |\epsilon_x| \leq 1; \text{ and}$$
$$\varphi''(x) = \varphi(x+2-\delta_x) - 2\varphi(x+1-\delta_x) + \varphi(x-\delta_x) \quad \text{for some } |\delta_x| \leq 2.$$

Thus, since $\varphi(x) \to 0$ as $|x| \to \infty$, it holds that $\varphi'(x)$ and $\varphi''(x)$ tend to zero as $|x| \to \infty$. Together with the fact that $\lambda'(x; b_m)$ and $\lambda''(x; b_m)$ are uniformly bounded over x and b_m, this implies that $R_2 \to 0$ as $l, m \to 0$.

The convergence of R_3 (toward zero) can also be shown by the uniform convergence of φ_l; (40) and (41); the uniform boundedness of $\lambda'(x; b_m)$ and $\lambda''(x; b_m)$; and the following inequality:

$$\begin{aligned} R_3 \leq & \sup_{x>1/(2b_m)} |\mathbf{M}_1(x)\mathbf{w}(x)| \times \sup_{x>1/(2b_m)} \left| \frac{d}{dx}\lambda(x; b_m)(\varphi(x) - \varphi_l(x)) \right| \\ & + \sup_{x>1/(2b_m)} |\mathbf{M}_2(x)\mathbf{w}(x)|/2 \times \sup_{x>1/(2b_m)} \left| \frac{d^2}{dx^2}\lambda(x; b_m)(\varphi(x) - \varphi_l(x)) \right|. \end{aligned}$$

This gives the desired result. ■

Proof of Lemma 3. (i) and (ii) can be shown in the same manner as (i), and thus we only prove (i).

$$||\mathcal{A}_0\phi_{l,m} - \mathcal{A}_0\eta_{l,m}||_{\mathbf{w}} \leq S_1 + S_2 + S_3$$

where

$$S_1 := ||\mathcal{A}_0\phi_{l,m} - \mathcal{A}_0\phi_{\infty,m}||_{\mathbf{w}}; \quad S_2 := ||\mathcal{A}_0\phi_{\infty,m} - \mathcal{A}_0\eta_{\infty,m}||_{\mathbf{w}};$$
$$S_3 := ||\mathcal{A}_0\eta_{\infty,m} - \mathcal{A}_0\eta_{l,m}||_{\mathbf{w}}.$$

$S_1 = R_3$ (R_3 is defined in Lemma 2), and thus $S_1 \to 0$. Next, look at S_2. By a change of variables and the interchangeability of differentiation and integration (which follows from the bounded convergence noting that all relevant functions are uniformly bounded),

$$\begin{aligned} S_2 \leq & \sup_{|x|>1/(2b_m)} \left| \mathcal{A}_0 \left(\lambda(x; b_m) \int_{-\infty}^{\infty} [\varphi(x) - \varphi(y)] \frac{1}{b_m}\rho\left(\frac{x-y}{b_m}\right) dy \right) \right| \mathbf{w}(x) \\ = & \sup_{|x|>1/(2b_m)} \left| \mathcal{A}_0 \left(\int_{-\infty}^{\infty} \lambda(x; b_m) [\varphi(x) - \varphi(x - pb_m)] \rho(p)\, dp \right) \right| \mathbf{w}(x) \\ \leq & \sup_{|x|>1/(2b_m)} |\mathbf{M}_1(x)\mathbf{w}(x)| \times \left\{ \sup_{|x|>1/(2b_m)} \left| \int_{-\infty}^{\infty} \lambda(x; b_m) [\varphi'(x) - \varphi'(x - pb_m)] \rho(p)\, dp \right| \right. \\ & \left. + \sup_{|x|>1/(2b_m)} \left| \int_{-\infty}^{\infty} \lambda'(x; b_m) [\varphi(x) - \varphi(x - pb_m)] \rho(p)\, dp \right| \right\} \\ & + \sup_{|x|>1/(2b_m)} |\mathbf{M}_2(x)\mathbf{w}(x)|/2 \times \sup_{|x|>1/(2b_m)} \left| \int_{-\infty}^{\infty} \frac{d^2}{dx^2} \{\lambda(x; b_m) [\varphi(x) - \varphi(x - pb_m)]\} \rho(p)\, dp \right|. \end{aligned}$$

Since the support of ρ is a compact interval $[u_1, u_2]$, it holds that

$$\sup_{|x|>1/(2b_m)} \left| \int_{-\infty}^{\infty} \lambda(x; b_m) \left[\varphi'(x) - \varphi'(x - pb_m)\right] \rho(p)\, dp \right|$$
$$\leq \sup_{\substack{|x|>1/(2b_m);\\ p\in[u_1,u_2]}} \left|\varphi'(x) - \varphi'(x - pb_m)\right| \times \left| \int_{-\infty}^{\infty} \rho(q)\, dq \right|$$
$$\leq 2 \sup_{|x|>1/(3b_m)} \left|\varphi'(x)\right| \to 0 \quad \text{as} \quad m \to \infty$$

where the first inequality holds by the uniform boundedness of $\lambda\,(\leq 1)$; and the second holds for large m. By a similar argument, it also holds that as $m \to 0$

$$\int_{-\infty}^{\infty} \lambda'(x; b_m) \left[\varphi(x) - \varphi(x - pb_m)\right] \rho(p)\, dp \to 0;$$
$$\sup_{|x|>1/(2b_m)} \left| \int_{-\infty}^{\infty} \frac{d^2}{dx^2} \left\{\lambda(x; b_m) \left[\varphi(x) - \varphi(x - pb_m)\right]\right\} \rho(p)\, dp \right| \to 0.$$

Therefore, we now have shown that $S_2 \to \infty$ as $m \to 0$.

By the same arguments as for S_2, we can obtain the bound of S_3 as follows:

$$S_3 \leq \sup_{|x|>1/(2b_m)} \left| \mathcal{A}_0 \left(\lambda(x; b_m) \int_{-\infty}^{\infty} \left[\varphi(y) - \varphi_l(y)\right] \frac{1}{b_m} \rho\left(\frac{x-y}{b_m}\right) dy \right) \right| \mathbf{w}(x)$$
$$\leq \sup_{|x|>1/(2b_m)} \left|\mathbf{M}_1(x)\, \mathbf{w}(x)\right|$$
$$\times \sup_{|x|>1/(2b_m)} \left| \int_{-\infty}^{\infty} \frac{d}{dx} \lambda(x; b_m) \left[\varphi(x - pb_m) - \varphi_l(x - pb_m)\right] \rho(p)\, dp \right|$$
$$+ \sup_{|x|>1/(2b_m)} \left|\mathbf{M}_2(x)\, \mathbf{w}(x)\right| /2$$
$$\times \sup_{|x|>1/(2b_m)} \left| \int_{-\infty}^{\infty} \frac{d^2}{dx^2} \lambda(x; b_m) \left[\varphi(x - pb_m) - \varphi_l(x - pb_m)\right] \rho(p)\, dp \right|.$$

Using (40) and (41) in Lemma 2, the right-hand side tends to zero as $l \to \infty$, completing the proof. ■

Proof of Lemma 1. Write $\varphi(x) = (\varphi(x)/\mathfrak{w}(x)) \times \mathfrak{w}(x) \equiv f(x) \times \mathfrak{w}(x)$. Let $w_c(x) := \exp\{-cx^2\}$ with $c > 0$, and

$$\hat{\mathbf{C}}^k(\mathbb{R}; w_c)$$
$$= \left\{ f\,(: \mathbb{R} \to \mathbb{R}) \mid f \text{ is } k\text{-times continuously differentiable; } f^{(i)} w_c \in \hat{\mathbf{C}}(\mathbb{R}) \text{ for } i = 0, \ldots, k \right\}.$$

By the assumption, $f(x)\,(= \varphi(x)/\mathfrak{w}(x)) \in \mathbf{C}_K^k(\mathbb{R})$. Since the support of f is compact, $f^{(i)}(x) \to 0$ as $|x| \to 0$ for any $i \leq k$. Thus, $f(x) \in \hat{\mathbf{C}}^k(\mathbb{R}; \mathfrak{w})$. By Lemma 4, there exists a sequence of function $\{L_J(x)\}$ such that each L_J is of degree at most $J-1$, and for arbitrary $c\,(>0)$

$$\sum_{i=0}^{k-1}\sup_{x\in\mathbb{R}}\left|f^{(i)}(x)-L_J^{(i)}(x)\right|w_c(x)\to 0 \text{ as } J\to\infty. \tag{42}$$

For $k=1$, (24) holds obviously by (42) by setting $c=1/2$. Consider the case when $k=2$:

$$\begin{aligned}\left|\varphi^{(1)}(x)-H_J^{(1)}(x)\right| &= \left|f^{(1)}(x)\mathfrak{w}(x)-f(x)x\mathfrak{w}(x)-L_J^{(1)}(x)\mathfrak{w}(x)-L_J(x)x\mathfrak{w}(x)\right| \\ &\le \left|f^{(1)}(x)-L_J^{(1)}(x)\right|\mathfrak{w}(x)+|f(x)-L_J(x)|\,|x\mathfrak{w}(x)| \\ &\le C\left|f^{(1)}(x)-L_J^{(1)}(x)\right|w_c(x)+C\,|f(x)-L_J(x)|\,w_c(x) \end{aligned} \tag{43}$$

where the last inequality holds since there exist some constants $C\,(>0)$ and $c\in(0,1/2)$, such that for any $i\in\{0,1,\ldots,k\}$,

$$\left|\left(d^i/dx^i\right)\mathfrak{w}(x)\right|\le Cw_c(x) \quad (C \text{ and } c \text{ depend only on } k). \tag{44}$$

By (42), the right-hand side of (43) towards 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 $k=1$, gives the desired result. When $k>2$, the proof can be done analogously (using (44) repeatedly). ■

Lemma 4 *Suppose that $f\in\hat{\mathbf{C}}^k(\mathbb{R};w_c)$ for some c. Then, there exists a sequence of polynomial functions $\{L_J(x): J=k+1,k+2,\ldots\}$ such that each L_J is of degree at most $J-1$, and*

$$\sum_{i=0}^{k-1}\sup_{x\in\mathbb{R}}\left|f^{(i)}(x)-L_J^{(i)}(x)\right|w_c(x)\to 0 \text{ as } J\to\infty. \tag{45}$$

Proof of Lemma 4. (i) By inequalities (2) and Corollary 1 of Balázs (2004), the left-hand side of (45) is bounded by

$$\sum_{i=0}^{k-1}\alpha_{i,k}\left(\frac{c_3J^{1/(1+c_4)}}{J+1-k}\right)^{k-i}\mathbf{E}_{J-k+1}\left(f^{(k)}\right)_{w_c}\log J, \tag{46}$$

where $\mathbf{E}_{J-k+1}(f)_{w_c}:=\inf_{p\in\Pi_J}\sup_{x\in\mathbb{R}}|w_c(x)(f(x)-p(x))|$; and Π_J denotes the set of polynomials of degree at most J. By the arguments in Szabados (1997), $\mathbf{E}_J(g)_{w_c}\to 0$ (as $J\to\infty$) for any continuous function g on $\mathbb{R}$ with $\lim_{|x|\to\infty}g(x)w_c(x)\to 0$, which implies that

$$\mathbf{E}_J\left(f^{(i)}\right)_{w_c}\to 0 \text{ for } i=0,1,\ldots,(k-1). \tag{47}$$

(46) and (47) gives the desired result. ■

Proof of Theorem 2. The "if" part is obvious. I prove the "only if" part. Suppose $\Pr\left[\mathcal{A}\varphi\left(X_s\right)=\mathcal{A}_0\varphi\left(X_s\right)\right]<1$. Then, there exist some set $U\left(\subset\mathfrak{B}\left(I\right)\right)$ such that $\Pr\left[X_s\in U\right]>0$ and

$$\forall X_s\in U,\quad\left(\mathcal{A}-\mathcal{A}_0\right)\varphi\left(X_s\right)\neq 0. \tag{48}$$

Let $\left\{H_J\left(x\right)\right\}$ be a sequence of functions given in Lemma 1 whose derivatives (upto any order) simultaneously and uniformly converge to $\varphi\left(x\right)$. Fix any x in the interior point of U. By Assumption 3, it also holds that for large J,

$$\left(\mathcal{A}-\mathcal{A}_0\right)H_J\left(x\right)\neq 0. \tag{49}$$

Recall that $H_J\left(x\right)=L_J\left(x\right)\mathfrak{w}\left(x\right)$ and L_J is a polynomial function of degree at most $\left(J-1\right)$. We can write

$$H_J\left(x\right)=\sum_{l=0}^{J}\gamma_l g_l\left(x\right)$$

where $g_l\left(x\right):=x^l\mathfrak{w}\left(x\right)$ and $\left\{\gamma_l\right\}$ is a sequence of (constant) coefficients. Using this expression, and the linearity of $\mathcal{A}$, $\mathcal{A}_0$, and the expectation operator, we have

$$\left(\mathcal{A}-\mathcal{A}_0\right)H_J\left(x\right)=\sum_{l=0}^{J}\gamma_l\left(\mathcal{A}-\mathcal{A}_0\right)g_l\left(x\right). \tag{50}$$

By (49) and (50), there exists some $l^*\left(\leq J\right)$ such that

$$\left(\mathcal{A}-\mathcal{A}_0\right)g_{l^*}\left(x\right)\neq 0, \tag{51}$$

which, together with the continuity of $\left(\mathcal{A}-\mathcal{A}_0\right)g_{l^*}\left(\cdot\right)$, implies the existence of some interval $\left[u_1,u_2\right]\left(\subset U\right)$ such that

$$\forall x\in\left[u_1,u_2\right],\quad\left(\mathcal{A}-\mathcal{A}_0\right)g_{l^*}\left(x\right)\neq 0. \tag{52}$$

It also holds that $\Pr\left[X_s\in\left[u_1,u_2\right]\right]>0$, by the maintained assumption of the existence of the invariant (marginal) density of X_s (the marginal distribution of X_s is absolutely continuous).

Now write

$$\eta\left(x;\xi\right)=\exp\left(\xi x\right)\mathfrak{w}\left(x\right)=\sum_{k=0}^{\infty}\frac{\xi^k}{k!}x^k\mathfrak{w}\left(x\right)=\lim_{J\to\infty}\eta_J\left(x;\xi\right)$$

where

$$\eta_J\left(x;\xi\right):=\sum_{l=0}^{J}\frac{\xi^l}{l!}x^l\mathfrak{w}\left(x\right)=\sum_{l=0}^{J}\frac{\xi^l}{l!}g_l\left(x\right).$$

By a similar argument as in Lemma 1, the simultaneous convergence of $\Lambda_J(\cdot;\xi)$ and its derivatives occurs uniformly, i.e., for any large $\bar{k}$, and for any ξ,

$$\sum_{i=0}^{\bar{k}} \sup_{x\in\mathbb{R}} \left|\eta^{(i)}(x;\xi) - \eta_J^{(i)}(x;\xi)\right| \to 0 \text{ as } J\to\infty. \tag{53}$$

From this, we have $\lim_{J\to\infty}\eta_J(x;\xi)$ well-defined and equal to $\lim_{J\to\infty}\eta_J(x;\xi) = \eta(x;\xi)$.

(53) allows the termwise differentiability of $\eta(x;\xi)$ (with respect to x), and therefore,

$$\begin{aligned}
\mathcal{A}_0\eta(x;\xi) &= \mu(x)\frac{d}{dx}\left(\sum_{l=0}^{\infty}\frac{\xi^l}{l!}g_l(x)\right) + \frac{\sigma^2(x)}{2}\frac{d^2}{dx^2}\left(\sum_{l=0}^{\infty}\frac{\xi^l}{l!}g_l(x)\right) \\
&= \mu(x)\left(\sum_{l=0}^{\infty}\frac{\xi^l}{l!}\frac{d}{dx}g_l(x)\right) + \frac{\sigma^2(x)}{2}\left(\sum_{l=0}^{\infty}\frac{\xi^l}{l!}\frac{d^2}{dx^2}g_l(x)\right) \\
&= \sum_{l=0}^{\infty}\frac{\xi^l}{l!}\mathcal{A}_0 g_l(x). \qquad (54)
\end{aligned}$$

Furthermore, by (53) and Assumption 3, we also have

$$\sum_{l=0}^{\infty}\frac{\xi^l}{l!}\mathcal{A}g_l(x) \tag{55}$$

well-defined and equal to $\mathcal{A}\eta(x;\xi)$ over any x for each ξ. (54) and (55) allow the termwise operation of $(\mathcal{A}-\mathcal{A}_0)$ to $\eta(x;\xi)$:

$$\begin{aligned}
(\mathcal{A}-\mathcal{A}_0)\eta(x;\xi) &= \sum_{l=0}^{\infty}\frac{\xi^l}{l!}\left\{\mathcal{A}g_l(x) - \mathcal{A}_0 g_l(x)\right\} \\
&\equiv L(\xi;x).
\end{aligned}$$

Regard $L(\xi;x)$ as a function of ξ for any $x\in[u_1,u_2]$. Then, $L(\xi;x)$ is a power series of ξ whose radius of convergence is ∞ (noting that the form of $\eta(x;\xi) = \exp(\xi x)\mathfrak{w}(x)$), and thus, termwise differentiation of $L(\cdot,x)$ is possible:

$$\begin{aligned}
\lim_{\xi\to 0}(d/d\xi)^{l^*}L(\xi;x) &= \lim_{\xi\to 0}\sum_{l=0}^{\infty}\left(\frac{d}{d\xi}\right)^{l^*}\left(\frac{\xi^l}{l!}(\mathcal{A}-\mathcal{A}_0)g_l(x)\right) \\
&= (\mathcal{A}-\mathcal{A}_0)g_{l^*}(x).
\end{aligned}$$

This and (52) imply that for any $x\in[u_1,u_2]$, there exists some ξ (in the neighborhood of zero) such that $L^2(\xi;x)>0$. The continuity of $L(\cdot;\cdot)$ implies the desired result. ■

Proof of Theorem 3. The "if" part is obvious, and thus, I prove the "only if" part. Suppose $\Pr[\mathcal{A}\varphi(X_s) = \mathcal{A}_0\varphi(X_s)] < 1$. Then by Lemma 6, Now, there exists some $\bar{\xi}_2$ (in the neighborhood of zero) and some subset $S_{\bar{\xi}_2}$ of $\left[0, \bar{\xi}_2\right]$ such that

$$\forall \xi_2 \in S_{\bar{\xi}_2}, \quad E\left[\left(\mathcal{A}-\mathcal{A}_0\right) \varphi\left(X_s\right) \times \eta\left(X_s ; \xi_2\right) q\left(X_s\right)\right] \neq 0. \tag{56}$$

Now, fix any $\xi_2 \left(\in S_{\bar{\xi}_2}\right)$. Let $\{H_J(x)\}$ be a sequence of functions given in Lemma 1 whose derivatives simulataneously and uniformly converges to $\varphi(x)$ for any x (take $\bar{k}$ large enough). By Assumption 3 , for large J, it also holds that

$$E\left[\left(\mathcal{A}-\mathcal{A}_0\right) H_J\left(X_s\right) \times \eta\left(X_s ; \xi_2\right) q\left(X_s\right)\right] \neq 0. \tag{57}$$

Recall that $H_J(x) = L_J(x)\,\mathfrak{w}(x)$ and $L_J(x)$ is a polynomial function of degree at most $(J-1)$. Therefore, we can write

$$H_J(x) = \sum_{l=0}^{J} \gamma_l g_l(x)$$

where $g_l(x) := x^l \mathfrak{w}(x)$ and $\{\gamma_l\}$ is a sequence of coefficients. Using this expression, and the linearlity of $\mathcal{A}$, $\mathcal{A}_0$, and the expectation operator, we have

$$\begin{aligned} & E\left[\left(\mathcal{A}-\mathcal{A}_0\right) H_J\left(X_s\right) \times \eta\left(X_s ; \xi_2\right) q\left(X_s\right)\right] \\ & =\sum_{l=0}^{J} \gamma_l E\left[\left(\mathcal{A}-\mathcal{A}_0\right) g_l\left(X_s\right) \times \eta\left(X_s ; \xi_2\right) q\left(X_s\right)\right]. \end{aligned} \tag{58}$$

(57) and (58) imply that there exists some l^* such that

$$E\left[\left(\mathcal{A}-\mathcal{A}_0\right) g_{l^*}\left(X_s\right) \times \eta\left(X_s ; \xi_2\right) q\left(X_s\right)\right] \neq 0. \tag{59}$$

Now, write

$$\begin{aligned} \eta\left(x ; \xi_1\right) & =\exp \left(\xi_1 x\right) \mathfrak{w}(x) \\ & =\sum_{k=0}^{\infty} \frac{\xi_1^k}{k!} x^k \mathfrak{w}(x)=\lim_{J \rightarrow \infty} \eta_J\left(x ; \xi_1\right) \end{aligned} \tag{60}$$

where

$$\eta_J\left(x ; \xi_1\right) := \sum_{l=0}^{J}\left(\xi_1^l / l!\right) x^l \mathfrak{w}(x)=\sum_{l=0}^{J}\left(\xi_1^l / l!\right) g_l(x)$$

Note that $\lim_{J \rightarrow \infty} \eta_J\left(x ; \xi_1\right)$ is well-defined and equal to $\lim_{J \rightarrow \infty} \eta_J\left(x ; \xi_1\right)=\eta\left(x ; \xi_1\right)$ since for any ξ_1 the simultaneous convergence of $\Lambda_J\left(x ; \xi_1\right)$ and its derivatives occurs uniformly over x, i.e., for any large $\bar{k}$

$$\sum_{i=0}^{\bar{k}} \sup_{x \in \mathbb{R}}\left|\eta^{(i)}\left(x ; \xi_1\right)-\eta_J^{(i)}(x)\right| \rightarrow 0 \quad \text{as} \quad J \rightarrow \infty, \tag{61}$$

which follows from a similar argument as in Lemma 1.

(61) allows termwise differentiability, and thus

$$\begin{aligned}\mathcal{A}_0\eta(x;\xi_1) &= \mu(x)\frac{d}{dx}\left(\sum_{l=0}^{\infty}\frac{\xi_1^l}{l!}g_l(x)\right)+\frac{\sigma^2(x)}{2}\frac{d^2}{dx^2}\left(\sum_{l=0}^{\infty}\frac{\xi_1^l}{l!}g_l(x)\right)\\ &= \mu(x)\left(\sum_{l=0}^{\infty}\frac{\xi_1^l}{l!}\frac{d}{dx}g_l(x)\right)+\frac{\sigma^2(x)}{2}\left(\sum_{l=0}^{\infty}\frac{\xi_1^l}{l!}\frac{d^2}{dx^2}g_l(x)\right)\\ &= \sum_{l=0}^{\infty}\frac{\xi_1^l}{l!}\mathcal{A}_0g_l(x). \end{aligned} \tag{62}$$

Furthermore, by (61) and Assumption 3, for each ξ_1

$$\sum_{l=0}^{\infty}\frac{\xi_1^l}{l!}\mathcal{A}g_l(x) \tag{63}$$

is well-defined over any x, and is equal to $\mathcal{A}\eta(x;\xi_1)$. Therefore, it follows from (62) and (63) that

$$\begin{aligned}E\left[(\mathcal{A}-\mathcal{A}_0)\eta(X_s;\xi_1)\times\Lambda(X_s;\xi_2)\right] &= \sum_{l=0}^{\infty}\frac{\xi_1^l}{l!}E\left[\{\mathcal{A}g_l(X_s)-\mathcal{A}_0g_l(X_s)\}\times\Lambda(X_s;\xi_2)\right]\\ &\equiv L(\xi_1;\xi_2). \end{aligned}$$

For each ξ_2, L is a power series of ξ_1 whose radius of convergence is any $\xi_1'\in(0,1)$, which allows the termwise differentiability of $L(\cdot,\xi_2)$:

$$\begin{aligned}\lim_{\xi_1\to 0}(d/d\xi_1)^{l^*}L(\xi_1,\xi_2) &= (d/d\xi_1)^{l^*}E\left[(\mathcal{A}-\mathcal{A}_0)\eta(X_s;\xi_1)\times\eta(X_s;\xi_2)q(X_s)\right]\\ &= \lim_{\xi_1\to 0}\left(\frac{d}{d\xi_1}\right)^{l^*}\left(\sum_{l=0}^{\infty}\frac{\xi_1^l}{l!}E\left[(\mathcal{A}-\mathcal{A}_0)g_l(x)\times\eta(X_s;\xi_2)q(X_s)\right]\right)\\ &= E\left[(\mathcal{A}-\mathcal{A}_0)g_{l^*}(x)\times\eta(X_s;\xi_2)q(X_s)\right]\neq 0. \end{aligned}$$

Therefore, there exists some ξ_1 (in the neighborhood of zero) such that

$$E\left[(\mathcal{A}-\mathcal{A}_0)\eta(X_s;\xi_1)\times\eta(X_s;\xi_2)q(X_s)\right]\neq 0.$$

which, together with the continuity of $L(\cdot,\cdot)$, implies the desired result. ■

Lemma 5 *Let Y be a random variable, $g(\cdot)$ $w_c(\cdot)$, and $q(\cdot)$ measurable functions: $w_c(x)=\exp\{-cx^2\}$ and $q(x)>0$ for any $x\in I$. Suppose that for some $c\in(0,1/2)$ $E[|g(Y)|w_c(Y)q(Y)]<\infty$. Then,* $\Pr[g(Y)=0]<1$ *if and only if there exists some $\zeta\in\mathbb{R}$ such that*

$$E\left[g(Y)\exp\{i\zeta Y\}w_c(Y)q(Y)\right]\neq 0.$$

Proof of Lemma 5. The 'if' part is obvious. I prove the 'only if ' part. Define

$$g_1(y) := \max\{g(y), 0\}; \quad g_2(y) := \max\{-g(y), 0\}; \quad \text{and}$$
$$b_1 := E[g_1(Y) w_c(Y) q(Y)]; \quad b_2 := E[g_2(Y) w_c(Y) q(Y)].$$

i) Consider the case that $b_1 > 0$ and $b_2 > 0$. Then, we can define probability measures for $j = 1, 2:$

$$v_j(B) := \int_B g_j(Y) w_c(Y) q(Y) v(dY) \Big/ b_j$$

where v is a probability mesure on $\mathbb{R}$ induced by Y; and B is an arbitrary Borel set in $\mathbb{R}$. We have

$$\begin{aligned}
&E[g(Y) \exp\{i\zeta Y\} w_c(Y)] \\
&= \int g_1(Y) \exp\{i\zeta Y\} w_c(Y) q(Y) v(dY) - \int g_2(Y) \exp\{i\zeta Y\} w_c(Y) q(Y) v(dY) \\
&= b_1 \int \exp\{i\zeta Y\} v_1(dY) - b_2 \int \exp\{i\zeta Y\} v_2(dY) \\
&= b_1 \eta_1(\xi) - b_2 \eta_2(\xi)
\end{aligned} \tag{64}$$

where $\eta_j(\zeta) := \int \exp\{i\zeta Y\} v_1(dY)$ is a characteristic function of v_j for $j = 1, 2$.

Suppose further that $E[g(Y) \exp\{i\zeta Y\} w_c(Y) q(Y)] = 0$ for any ζ. Then,

$$b_1 \eta_1(\xi) - b_2 \eta_2(\xi) = 0 \text{ for any } \zeta,$$

which implies that $b_1 = b_2$. Hence, by (64),

$$\eta_1(\zeta) = \eta_2(\zeta) \quad \text{for any } \zeta.$$

This is equivalent to

$$v_1(B) = v_2(B) \quad \text{for all } B.$$

We have $\int_B g(Y) v(dY) = 0$ for all B. Let $B_1 := \{Y : g_1(Y) > 0\}$. Since B_1 is a Borel set,

$$\int_{B_1} g(Y) v(dY) = 0,$$

which is only possible if B_1 is a null set with respect to v. Similarly, $B_2 := \{Y : g_1(Y) < 0\}$ is also a null set. Thus, we have $v(B_1 \cup B_2) = 0$ or

$$\Pr[g(Y) \neq 0] = 0.$$

We have shown that $\Pr[g(Y) = 0] < 1$ implies $E[g(Y) \exp\{i\zeta Y\} w_c(Y)] \neq 0$ for some r, given $b_1 > 0$ and $b_2 > 0$.

ii) Assume $b_1 > 0$ and $b_2 = 0$. Then,

$$E\left[g\left(Y\right)\exp\left\{i\zeta Y\right\}w_c\left(Y\right)q\left(Y\right)\right] = \int g_1\left(Y\right)\exp\left\{i\zeta Y\right\}w_c\left(Y\right)q\left(Y\right)v\left(dY\right)$$
$$= b_1\int \exp\left\{i\zeta Y\right\}v_1\left(dY\right) = b_1\eta_1\left(r\right)$$

where v_1 and η_1 are defined as above. By a similar argument as above, we can show that "$E\left[g\left(Y\right)\exp\left\{i\zeta Y\right\}w_c\left(Y\right)q\left(Y\right)\right] = 0$ for any ζ" implies that $\Pr\left[g_1\left(Y\right)\neq 0\right] = 0$. Noting that $b_2 = 0$ means $\Pr\left[g_2\left(Y\right)=0\right] = 1$, we have shown $\Pr\left[g\left(Y\right)=0\right] < 1$ implies $E\left[g\left(Y\right)\exp\left\{i\zeta Y\right\}w_c\left(Y\right)q\left(Y\right)\right]\neq 0$ for some r, given $b_1 > 0$ and $b_2 = 0$.

iii) When $b_1 = 0$ and $b_2 > 0$, we can apply the same argument. iv) When $b_1 = b_2 = 0$, it necessarily holds that $\Pr\left[g\left(Y\right)=0\right] = 1$. These arguments give the desired result. ■

Lemma 6 *Under the same notations and conditions as in Lemma 5,*

$$\Pr\left[g\left(Y\right)=0\right] < 1$$

if and only if there exist some $\bar{\zeta} > 0$ (arbitrarily close to zero) and some subset $S_{\bar{\zeta}}$ of $\left[-\bar{\zeta},\bar{\zeta}\right]$ (with the Lebesgue measure of $S_{\bar{\zeta}}$ non-zero) such that

$$\forall\zeta\in S_{\bar{\zeta}},\quad E\left[g\left(Y\right)\exp\left(\zeta Y\right)\mathfrak{w}\left(Y\right)q\left(Y\right)\right]\neq 0.$$

Proof of Lemma 6. The "if" part is obvious. I prove the "only if" part. Suppose $\Pr\left[g\left(Y\right)=0\right] < 1$. By Lemma 5, for any (fixed) ζ,

$$E\left[g\left(Y\right)\exp\left\{i\zeta Y\right\}\mathfrak{w}\left(Y\right)q\left(Y\right)\right] = E\left[g\left(Y\right)\sum_{l=0}^{\infty}\frac{\left(i\zeta\right)^l}{l!}Y^l\mathfrak{w}\left(Y\right)q\left(Y\right)\right]$$
$$= \sum_{k=0}^{\infty}E\left[g\left(Y\right)\frac{\left(i\zeta\right)^l}{l!}Y^l\mathfrak{w}\left(Y\right)q\left(Y\right)\right]\neq 0. \tag{65}$$

We obtain the second equality by applying the dominant convergence theorem: the dominant function exists since $E\left[\left|g\left(Y\right)\right|w_c\left(x\right)q\left(Y\right)\right] < \infty$ and

$$\forall J,\quad \sum_{k=0}^{J}\left|\frac{\zeta^l}{l!}Y^l\right|\mathfrak{w}\left(Y\right)q\left(Y\right)\leq\frac{1}{\sqrt{2\pi}}\left|g\left(Y\right)\right|w_c\left(x\right)q\left(Y\right)\underbrace{\sum_{l=0}^{\infty}\left|\frac{\zeta^l}{l!}Y^lw_d\left(Y\right)\right|}_{<\infty}$$

where $d := \left(1/2-c\right)\in\left(0,1/2\right)$, and $\mathfrak{w}\left(x\right) = \left(1/\sqrt{2\pi}\right)w_c\left(x\right)w_d\left(x\right)$. It holds that $\sum_{l=0}^{\infty}\left|Y^lw_d\left(Y\right)\right| < \infty$ by the following arguments: first, the roots of $\left(d/dy\right)\left[x^lw_d\left(x\right)\right] = 0$ are $y = 0, \pm\sqrt{l/2d}$, and thus

$$\left|Y^lw_d\left(Y\right)\right|\leq\left(\frac{l}{2d}\right)^{l/2}\exp\left\{-l/2d\right\}. \tag{66}$$

Second, by (66) and Starling's formula: $l!/\sqrt{2\pi l}l^l \exp\{-l\} \to 1$ as $l \to \infty$, there exist some $m > 2$ and a positive constant C (which depends on m) such that

$$\begin{aligned}
&\sum_{l=0}^{\infty} \frac{|\zeta|^l}{l!} \left|Y^{l+1} w_d(Y)\right| \\
&= \text{constant} + \sum_{l=1}^{\infty} \frac{|\zeta|^l}{\sqrt{2\pi l}l^l \exp\{-l\}} \left(\frac{l}{1-2c}\right)^{l/2} \exp\{-l/2(1-2c)\} \times \frac{\sqrt{2\pi l}l^l \exp\{-l\}}{l!} \\
&\leq \text{constant} + \sum_{l=1}^{\infty} C l^{-m}. \qquad (67)
\end{aligned}$$

Finally, by (66) and (67),

$$\sum_{l=0}^{\infty} \left|\frac{Y^l}{l!} w_d(Y)\right| < \infty.$$

Now, (65) implies that there exists some l^* such that

$$E\left[g(Y) Y^{l^*} w(Y) q(Y)\right] \neq 0. \qquad (68)$$

Consider the derivative:

$$\begin{aligned}
&\left(\frac{d}{d\zeta}\right)^{l^*} E\left[g(Y)\exp(\zeta X_s)\mathfrak{w}(Y) q(Y)\right] \\
&= \frac{d^{l^*}}{d\zeta^{l^*}} E\left[g(Y) \sum_{l=0}^{\infty} \frac{\zeta^l}{l!} Y^l \mathfrak{w}(Y) q(Y)\right] \\
&= \frac{d^{l^*}}{d\zeta^{l^*}} \left(\sum_{l=0}^{\infty} \frac{\zeta^l}{l!} E\left[g(Y) Y^l \mathfrak{w}(Y) q(Y)\right]\right) = \sum_{l=0}^{\infty} \frac{d^{l^*}}{d\zeta^{l^*}} \left(\frac{\zeta^l}{l!} E\left[g(Y) Y^l \mathfrak{w}(Y) q(Y)\right]\right) \qquad (69)
\end{aligned}$$

where the second equality holds by the same arguments as above, and the third equality holds by the termwise differentiability of a power series within the radius of convergence, noting that $\sum_{l=0}^{J} \frac{\zeta^l}{l!} E\left[g(Y) Y^l \mathfrak{w}(Y) q(Y)\right]$ converges to $\sum_{l=0}^{\infty} \frac{\zeta^l}{l!} E\left[g(Y) Y^l \mathfrak{w}(Y) q(Y)\right]$ uniformly for any $|\zeta| \leq \bar{\zeta}$ ($\bar{\zeta} \in (0, \infty)$ can be arbitrarily). Letting $\zeta \to 0$, the right-hand side of (69) approaches $E\left[g(Y) Y^l w(Y) q(Y)\right]$, which together with (68) implies that

$$\lim_{\zeta \to 0} \frac{d^l}{d\zeta^l} E\left[g(Y)\exp(\zeta X_s)\mathfrak{w}(Y) q(Y)\right] \neq 0.$$

Thus $E\left[g(Y)\exp(\zeta X_s)\mathfrak{w}(Y) q(Y)\right] \neq 0$ for some ζ in the neighborhood of 0. This and the continuity of $E\left[g(Y)\exp(\zeta X_s)\mathfrak{w}(Y) q(Y)\right]$ (with respect to ζ) imply the existence of the set $S_{\bar{\zeta}}$, completing the proof. ■

B Proofs of Theorems 4 and 5

Proof of Theorem 4. Let $\hat{F}$ be the empirical distribution function of $\{X_{j\Delta}\}$ and $F(x)$ the marginal distribution function $\{X_s\}$, i.e., $F(x) = \int_{-\infty}^{x} \pi(x)\, dx$. Look at

$$\begin{aligned}
\hat{N}(\xi) &= \int \hat{\lambda}^2(x;\xi)\, d\hat{F}(x) \\
&= \int \left\{ \frac{1}{Th} \sum_{j=1}^{n-1} K\left(\frac{X_{j\Delta}-x}{h}\right) \eta'''(X_{j\Delta};\xi)\, \sigma^3(X_{j\Delta}) \int_{j\Delta}^{(j+1)\Delta} \int_{j\Delta}^{s} \int_{j\Delta}^{u} dW_v dW_u dW_s \right\}^2 dF(x) \\
&\quad + o_p\left(\frac{T}{n^2 h^{1/2}}\right) \\
&= 2U_n(\xi) + B_n(\xi) + o_p\left(\frac{T}{n^2 h^{1/2}}\right)
\end{aligned}$$

where the equalities hold uniformly over ξ; the second holds by Lemma 7; and

$$\begin{aligned}
U_n(\xi) &:= \left(\frac{1}{Th}\right)^2 \sum_{1\le j<k\le(n-1)} \int \Psi_\xi(j\Delta;x)\, \Psi_\xi(k\Delta;x)\, dF(x); \\
B_n(\xi) &:= \left(\frac{1}{Th}\right)^2 \sum_{j=1}^{n-1} \int \Psi_\xi^2(j\Delta;x)\, dF(x); \text{ and} \\
\Psi_\xi(j\Delta;x) &:= K\left(\frac{X_{j\Delta}-x}{h}\right) \eta'''(X_{j\Delta};\xi)\, \sigma^3(X_{j\Delta}) \int_{j\Delta}^{(j+1)\Delta} \int_{j\Delta}^{s} \int_{j\Delta}^{u} dW_v dW_s dW_s.
\end{aligned}$$

U_n is in the form of a U-statistic, and B_n is the bias term. Lemma 8, 9, and 10 show that (i) finite dimentional distributions of $U_n(\xi)/\sqrt{T^2/n^4 h}$ is characterized as zero-mean normal with covariance kernel Λ_N; (ii) the tightness of $U_n(\xi)/\sqrt{T^2/n^4 h}$; and (iii) the uniform convergence of the normalized $B_n(\xi)$. Given (i)-(iii), we can obtain the weak convergence result as desired (see, e.g., Billingsley (1999)). ■

Lemma 7 *Under the same conditions as Theorem 4,*

$$\begin{aligned}
&\int \hat{\lambda}^2(x;\xi)\, d\hat{F}(x) \\
&= \int \left\{ \frac{1}{Th} \sum_{j=1}^{n-1} K\left(\frac{X_{j\Delta}-x}{h}\right) \eta'''(X_{j\Delta};\xi)\, \sigma^3(X_{j\Delta}) \int_{j\Delta}^{(j+1)\Delta} \int_{j\Delta}^{s} \int_{j\Delta}^{u} dW_v dW_u dW_s \right\}^2 dF(x) \\
&+ o_p\left(\frac{T}{n^2 h^{1/2}}\right)
\end{aligned}$$

where the equality holds uniformly over ξ.

Proof of Lemma 7. By Ito's lemma,

$$\eta^{(k)}\left(X_{(j+1)\Delta};\xi\right)-\eta\left(X_{j\Delta};\xi\right)=\int_{j\Delta}^{(j+1)\Delta}\left[\eta^{(k+1)}\left(X_s;\xi\right)\mu\left(X_s\right)+\frac{1}{2}\eta^{(k+2)}\left(X_s;\xi\right)\sigma^2\left(X_s\right)\right]ds$$
$$+\int_{j\Delta}^{(j+1)\Delta}\eta^{(k+1)}\left(X_s;\xi\right)\sigma\left(X_s\right)dW_s\quad\text{for}\quad k=0,1,2; \tag{70}$$

and

$$\left[X_{(j+1)\Delta}-X_{j\Delta}\right]^2=\int_{j\Delta}^{(j+1)\Delta}\left\{2\left[X_{(j+1)\Delta}-X_{j\Delta}\right]\mu\left(X_s\right)+\sigma^2\left(X_s\right)\right\}ds$$
$$+\int_{j\Delta}^{(j+1)\Delta}2\left[X_{(j+1)\Delta}-X_{j\Delta}\right]\sigma\left(X_s\right)dW_s. \tag{71}$$

By (70) and (71), we can show

$$\begin{aligned}&\hat{\lambda}\left(x;\xi\right)\\&=O_p(\Delta^{3/2}\sqrt{\Delta\log\left(1/\Delta\right)})\times\frac{1}{Th}\sum_{j=1}^{n-1}K\left(\frac{X_{j\Delta}-x}{h}\right)\\&+\frac{1}{Th}\sum_{j=1}^{n-1}K\left(\frac{X_{j\Delta}-x}{h}\right)\eta'''\left(X_{j\Delta};\xi\right)\sigma^3\left(X_{j\Delta}\right)\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\int_{j\Delta}^{u}dW_sdW_udW_s\\&\text{uniformly over }\xi,\end{aligned}$$

where we use the infill assumption and the modulus of continuity of the diffusion (regard $\eta^{(k)}$ as the *damping* function in the sense of Kanaya (2007). The uniformity with respect to ξ follows from the compactness of Ξ.

Next, look at

$$\begin{aligned}\int\hat{\lambda}^2\left(x;\xi\right)d\hat{F}\left(x\right)&=\int\hat{\lambda}^2\left(x;\xi\right)dF\left(x\right)+\int\hat{\lambda}^2\left(x;\xi\right)d\left[\hat{F}\left(x\right)-F\left(x\right)\right]\\&\equiv D_{1,n}+D_{2,n}.\end{aligned}$$

We first analyze $D_{1,n}$. Look at

$$\begin{aligned}D_{1,n}&=D_{1-1,n}+D_{1-2,n}\\&+\int\left\{\frac{1}{Th}\sum_{j=1}^{n-1}K\left(\frac{X_{j\Delta}-x}{h}\right)\times\eta'''\left(X_{j\Delta};\xi\right)\sigma^3\left(X_{j\Delta}\right)\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\int_{j\Delta}^{u}dW_sdW_udW_s\right\}^2dF\left(x\right)\end{aligned}$$

where

$$D_{1-1,n} := O_p(\Delta^4 \log(1/\Delta)) \times \frac{1}{T^2h^2} \sum_{1 \le j,k \le n-1} \int K\left(\frac{X_{j\Delta} - x}{h}\right) K\left(\frac{X_{k\Delta} - x}{h}\right) dF(x);$$

$$D_{1-2,n} := O_p(\Delta^{3/2}\sqrt{\Delta \log(1/\Delta)}) \times \frac{2}{T^2h^2} \sum_{1 \le j,k \le n-1} \int K\left(\frac{X_{j\Delta} - x}{h}\right) K\left(\frac{X_{k\Delta} - x}{h}\right) dF(x)$$

$$\times \eta'''(X_{k\Delta};\xi)\,\sigma^3(X_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} \int_{k\Delta}^{s} \int_{k\Delta}^{u} dW_s dW_u dW_s.$$

By the same arguments as in Lemma A.2 of Gao and King (2004),

$$\frac{\Delta^2}{T^2h^2} E\left[\left| \sum_{1 \le j,k \le n-1} \int K\left(\frac{X_{j\Delta} - x}{h}\right) K\left(\frac{X_{k\Delta} - x}{h}\right) dF(x) \right|\right] = O(h),$$

and thus $D_{1-1,n} = O_p(\Delta^2 \log(1/\Delta)\,h)$. In a similar way, we can also show that $D_{1-1,n} = O_p(\Delta^2\sqrt{\log(1/\Delta)}h)$. $D_{2,n}$ can be analyzed using similar ways to Lemma 3 of Li (2007), and we have

$$D_{2,n} = o_p(\Delta^2 \log(1/\Delta)\,h).$$

Under the stated conditions, we have the desired result. ■

Lemma 8 *Under the same conditions as Theorem 4, finite dimensional distributions of*

$$\frac{U_n}{\sqrt{T^2/n^4 h}}$$

are characterized as mean-zero normal with covariance kernel

$$\Lambda_N(\xi,\xi') = \frac{1}{72}\int L^2(w)\,dw \int \left[\eta'''(y;\xi)\,\eta'''(y;\xi')\,\sigma^6(y)\,\pi^2(y)\right]^2 dy.$$

Proof. Define $Z_m(\xi)$ as

$$Z_m(\xi) := \sum_{j=1}^{m-1} \int \Psi_\xi(j\Delta;x)\,\Psi_\xi(m\Delta;x)\,dF(x) \quad \text{for} \quad m = 3,\dots,n.$$

Then, U_n can be rewritten as a sum of $\{Z_m\}$:

$$U_n(\xi) = \left(\frac{1}{Th}\right)^2 \sum_{m=3}^{n} Z_m(\xi).$$

Note that $Z_m(\xi)$ is $\mathfrak{F}_{m\Delta}$-measurable and $\{Z_m(\xi)\}_{m=3}^{n}$ is a martingale difference sequence (for each ξ). Applying the central limit theorem (with the aid of the Cramér-Wold device), we find the finite dimensional distribution of the normalized version of $U_n(\xi)$ (required conditions for the CLT can be checked, using the uniform boundedness of relevant functions). The covariance kernel can be computed as follows. First, observe that

$$\begin{aligned}
&\sum_{m=3}^{n} E\left[Z_m(\xi) Z_m(\xi')\right] \\
&= \sum_{m=3}^{n} E\left[\left(\sum_{j=1}^{m-1}\int \Psi_{\xi}(j\Delta; x)\,\Psi_{\xi}(m\Delta; x)\,dF(x)\right)\left(\sum_{j=1}^{m-1}\int \Psi_{\xi'}(j\Delta; y)\,\Psi_{\xi'}(m\Delta; y)\,dF(y)\right)\right] \\
&= A_n(\xi,\xi') + \sum_{2\le j<k<l\le n} E\left[\Pi_{\xi}(j,l)\,\Pi_{\xi'}(k,l)\right]
\end{aligned}$$

where

$$\begin{aligned}
A_n(\xi,\xi') &:= \sum_{2\le j<k\le n} E\left[\left(\int \Psi_{\xi}(j\Delta; x)\,\Psi_{\xi}(k\Delta; x)\,dF(x)\right)\left(\int \Psi_{\xi'}(j\Delta; y)\,\Psi_{\xi'}(k\Delta; y)\,dF(y)\right)\right]; \\
\Pi_{\xi}(j,k) &:= \int \Psi_{\xi}(j\Delta; x)\,\Psi_{\xi}(k\Delta; x)\,dF(x).
\end{aligned}$$

Using the mixing property of the process we can show

$$\sum_{2\le j<k<l\le n} E\left[\Pi_{\xi}(j,l)\,\Pi_{\xi'}(k,l)\right] = o_p(1),$$

in a similar manner to Gao and King (2004). Letting

$$\begin{aligned}
A_n(\xi,\xi') &= \int\int E\left[\sum_{2\le j<k\le n}\left[\Psi_{\xi}(j\Delta; x)\,\Psi_{\xi'}(j\Delta; y)\right]\times\left[\Psi_{\xi}(k\Delta; x)\,\Psi_{\xi'}(k\Delta; y)\right]\right] dF(x)\,dF(y) \\
&= A_{n,1}(\xi,\xi')/2 - A_{n,2}(\xi,\xi')/2
\end{aligned}$$

where

$$\begin{aligned}
A_{n,1}(\xi,\xi') &:= \int\int E\left[\left(\sum_{j=2}^{n}\Psi_{\xi}(j\Delta; x)\,\Psi_{\xi'}(j\Delta; y)\right)^2\right] dF(x)\,dF(y); \\
A_{n,2}(\xi,\xi') &:= \int\int E\left[\sum_{j=2}^{n}\Psi_{\xi}^2(j\Delta; x)\,\Psi_{\xi'}^2(j\Delta; y)\right] dF(x)\,dF(y),
\end{aligned}$$

we can also show

$$A_{n,2}(\xi,\xi') = o_p(1) \quad \text{uniformly over } \xi.$$

Now, to find the limit of $A_{n,1}(\xi,\xi')$, let

$$Y_n(x,y,\xi,\xi') := \sum_{j=2}^{n}\Psi_{\xi}(j\Delta; x)\,\Psi_{\xi'}(j\Delta; y)$$

and look at

$$Y_n(x,y,\xi,\xi') \xrightarrow{p} \frac{n\Delta^3}{6}\int K\left(\frac{p-x}{h}\right)K\left(\frac{p-y}{h}\right)\eta'''(p;\xi)\,\eta'''(p;\xi')\,\sigma^6(p)\,\pi(p)\,dp, \tag{72}$$

which follows from the LLN and

$$\begin{aligned}
&E\left[Y_n(x,y,\xi,\xi')\right] \\
&= E\left[\sum_{j=2}^{n} K\left(\frac{X_{j\Delta}-x}{h}\right)K\left(\frac{X_{j\Delta}-y}{h}\right)\right. \\
&\left.\times\eta'''(X_{j\Delta};\xi)\,\eta'''(X_{j\Delta};\xi')\,\sigma^6(X_{j\Delta})\left(\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\int_{j\Delta}^{u} dW_v dW_s dW_s\right)^2\right] \\
&= (n-2)\,\frac{\Delta^3}{6}\int K\left(\frac{p-x}{h}\right)K\left(\frac{p-y}{h}\right)\eta'''(p;\xi)\,\eta'''(p;\xi')\,\sigma^6(p)\,\pi(p)\,dp.
\end{aligned}$$

The sequence $\{Y_n(x,y,\xi,\xi')\}_{n\geq 1}$ is shown to be uniformly integrable, and thus, by generalized dominated convergence,

$$A_{n,1}(\xi,\xi') = \int\int E\left[Y_n^2(x,y,\xi,\xi')\right] dF(x)\,dF(y)$$

converges to

$$\begin{aligned}
&\left(\frac{n\Delta^3}{6}\right)^2 \int\int\int\int K\left(\frac{p-x}{h}\right)K\left(\frac{p-y}{h}\right)\eta'''(p;\xi)\,\eta'''(p;\xi')\,\sigma^6(p)\,\pi(p) \\
&\times K\left(\frac{q-x}{h}\right)K\left(\frac{q-y}{h}\right)\eta'''(q;\xi)\,\eta'''(q;\xi')\,\sigma^6(q)\,\pi(q)\,\pi(x)\,\pi(y)\,dpdqdxdy \\
&= \left(\frac{n\Delta^3}{6}\right)^2 \int\int\int\int K\left(\frac{p-x}{h}\right)K\left(\frac{p-x}{h}-\frac{x-y}{h}\right)\eta'''(p;\xi)\,\eta'''(p;\xi')\,\sigma^6(p)\,\pi(p) \\
&\times K\left(\frac{q-y}{h}-\frac{x-y}{h}\right)K\left(\frac{q-y}{h}\right)\eta'''(q;\xi)\,\eta'''(q;\xi')\,\sigma^6(q)\,\pi(q) \\
&\times \pi(x)\,\pi(y)\,dpdqdxdy \\
&= \left(\frac{n\Delta^3}{6}\right)^2 h^3 \int\int\int\int K(u)\,K(u-w)\,\eta'''(uh+wh+y;\xi)\,\eta'''(uh+wh+y;\xi') \\
&\times \sigma^6(uh+wh+y)\,\pi(uh+wh+y) \\
&\times K(v-w)\,K(v)\,\eta'''(vh+y;\xi)\,\eta'''(vh+y;\xi')\,\sigma^6(vh+y)\,\pi(vh+y) \\
&\times \pi(wh+y)\,\pi(y)\,dudvdwdy
\end{aligned}$$

$$= \left(\frac{n\Delta^3}{6}\right)^2 h^3 \int\int\int\int K(u) K(u-w) \eta'''(y;\xi) \eta'''(y;\xi') \sigma^6(y) \pi(y)$$
$$\times K(v-w) K(v) \eta'''(y;\xi) \eta'''(y;\xi') \sigma^6(y) \pi(y)$$
$$\times \pi(y) \pi(y) \, dudvdwdy$$
$$= \frac{n^2\Delta^6 h^3}{36} \int L^2(w) \, dw \int \left[\eta'''(y;\xi) \eta'''(y;\xi') \sigma^6(y) \pi^2(y)\right]^2 dy \, [1 + O(h)] .$$

From these results, we obtain

$$\sum_{m=3}^{n} E\left[Z_m(\xi) Z_m(\xi')\right] = A_{n,1}(\xi, \xi') / 2 + o(1)$$
$$= n^2 \Delta^6 h^3 \Lambda_N(\xi, \xi') + o(1) .$$

■

Lemma 9 *Under the same conditions as Theorem 4,* $U_n(\xi) / \sqrt{T^2/n^4 h}$ *is tight.*

Proof of Lemma 9. According to Theorem 7.3 of Billingsley (1999), it suffices to prove: (i) for each $\delta > 0$ and an arbitrary $\xi \in \Xi$, there exists an ε such that

$$\sup_n \Pr\left(U_n(\xi) / \sqrt{T^2/n^4 h} > \varepsilon\right) \leq \delta;$$

(ii) for each $\delta > 0$ and $\varepsilon > 0$ there exists an $\rho > 0$ such that

$$\sup_n \Pr\left(\sup_{|\xi - \xi'| < \rho} |U_n(\xi) - U_n(\xi')| / \sqrt{T^2/n^4 h} \geq \varepsilon\right) \leq \delta.$$

Condition (i) follows from the pointwise convergence result of $U_n(\xi) / \sqrt{T^2/n^4 h}$ in Lemma 8. Condition (ii) can be proved by the uniform Hölder property of $\eta(x;\xi)$ with respect to ξ. ■

Lemma 10 *Under the same conditions as Theorem 4,*

$$B_n(\xi) = \left(\Delta^2/Th\right) E_N(\xi) + o_p\left(\Delta^2/Th\right)$$

uniformly over ξ.

Proof of Lemma 10. The desired result follows from the stationarity and ergodicity (use the LLN), and

$$
\begin{aligned}
&E\left[B_n(\xi)\right] \\
&= \frac{n-1}{T^2h^2}E\left[\int K^2\left(\frac{X_{j\Delta}-x}{h}\right)dF(x)\left[\eta'''(X_{j\Delta};\xi)\sigma^3(X_{j\Delta})\right]^2\left(\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\int_{j\Delta}^{u}dW_v dW_s dW_s\right)^2\right] \\
&= \frac{n-1}{T^2h^2}\frac{\Delta^3}{6}\int\int K^2\left(\frac{p-x}{h}\right)\left[\eta'''(p;\xi)\sigma^3(p)\right]^2\pi(x)\pi(p)\,dxdp \\
&= \frac{n-1}{T^2h^2}\frac{\Delta^3}{6}\int\int K^2\left(\frac{p-x}{h}\right)\left[\eta'''(p;\xi)\sigma^3(p)\right]^2\pi(x)\pi(p)\,dxdp \\
&= \frac{n-1}{T^2h}\frac{\Delta^3}{6}\int\int K^2(q)\left[\eta'''(qh+x;\xi)\sigma^3(qh+x)\right]^2\pi(x)\pi(qh+x)\,dxdq \\
&= \frac{\Delta^2}{Th}\times\frac{1}{6}\int\int K^2(q)\,dq\times\int\left[\eta'''(x;\xi)\sigma^3(x)\right]^2\pi^2(x)\,dx\times\left[1+O(h)\right]
\end{aligned}
$$

where the last equality holds since $g(y) := \left[\eta'''(y;\xi)\sigma^3(y)\right]^2\pi(y)$ is globally Hölder (over y) with the coefficient uniform over ξ. ■

Proof of Theorem 5. Let $\hat{F}$ be the empirical distribution function of $\{X_{j\Delta}\}$ and $F(x)$ the marginal distribution function $\{X_s\}$, i.e., $F(x) = \int^x \pi(x)\,dx$. Letting

$$
\begin{aligned}
Y_{j,n}(\xi_1,\xi_2) &:= \frac{1}{Th}\int K\left(\frac{X_{j\Delta}-x}{h}\right)\eta(x;\xi_2)\pi(x)\,dF(x) \\
&\quad\times\eta'''(X_{j\Delta};\xi_1)\sigma^3(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\int_{j\Delta}^{u}dW_v dW_u dW_s \quad (j\le n),
\end{aligned}
$$

we can show that

$$
\hat{N}_2(\xi_1,\xi_2) = \sum_{j=1}^{n-1}Y_{j,n}(\xi_1,\xi_2) + o_p\left(T^{1/2}/n\right) \tag{73}
$$

holds uniformly over any $(\xi_1,\xi_2)\in\Xi_1\times\Xi_2$ by similar techniques found in Lemma 7 of Aït-Sahalia, Bickel and Stoker (1998) and Kanaya (2007) (regard η and its derivatives as "damping" functions (in the sense of Kanaya (2007)), and use Ito's lemma and path-continuity of diffusion processes). The uniformity follows from the global Hölder property of $\eta(x;\xi)$ and its derivatives (with respect to ξ).

Note that $\{Y_{j,n}(\xi_1,\xi_2)\}$ is a martingale-difference array. Applying the central limit theorem (with the aid of the Cramér-Wold device), we can show that finite-dimensional distributions of $\hat{Z}_2(\xi_1,\xi_2) = \left(n/T^{1/2}\right)\hat{N}(\xi_1,\xi_2)$ converges to those of $Z_{2,0}(\xi_1,\xi_2)$. Required conditions for the CLT, i.e.,

$$
\frac{T}{n^2}\sum_{j=1}^{n-1}E\left[Y_{j,n}(\xi_1,\xi_2)\,Y_{j,n}(\xi_1',\xi_2')\;\middle|\;\mathfrak{F}_{j\Delta}\right]\xrightarrow{p}\Lambda_{N_2}\left((\xi_1,\xi_2),(\xi_1',\xi_2')\right), \tag{74}
$$

and the Lindberg condition, are satisfied by the uniform boundedness of relevant functions $\eta(x;\xi_2)\pi^2(x)$ and $\eta'''(x;\xi_1)\sigma^3(x)$. Stochastic equicontinuity of $\hat{Z}_2(\xi_1,\xi_2)$ also follows from the global Hölder property of $\eta(x;\xi)$ and its derivatives. We have established the weak convergence of $\hat{Z}_2(\xi_1,\xi_2)$. The second part of the theorem follows from the continuous mapping theorem.

Proof of (74) (74) follows from the stationarity and ergodicity of the process, and

$$\begin{aligned}
&E\left[Y_{j,n}(\xi_1,\xi_2)\,Y_{j,n}(\xi_1',\xi_2')\right]\\
&=E\Bigg[\int K\left(\frac{X_{j\Delta}-x}{h}\right)\eta(x;\xi_2)\,\pi(x)\,F(x)\times\int K\left(\frac{X_{j\Delta}-y}{h}\right)\eta(y;\xi_2')\,\pi(y)\,dF(y)\\
&+\eta'''(X_{j\Delta};\xi_1)\,\eta'''(X_{j\Delta};\xi_1')\,\sigma^6(X_{j\Delta})\left(\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\int_{j\Delta}^{u}dW_v dW_u dW_s\right)^2\Bigg]\\
&=\frac{\Delta^3}{6}E\Bigg[\int K\left(\frac{X_{j\Delta}-x}{h}\right)\eta(x;\xi_2)\,\pi^2(x)\,dx\times\int K\left(\frac{X_{j\Delta}-y}{h}\right)\eta(y;\xi_2')\,\pi^2(y)\,dy\\
&\quad+\eta'''(X_{j\Delta};\xi_1)\,\eta'''(X_{j\Delta};\xi_1')\,\sigma^6(X_{j\Delta})\Bigg]\\
&=\frac{\Delta^3}{6}\int\int\int K\left(\frac{z-x}{h}\right)\eta(x;\xi_2)\,\pi^2(x)\,K\left(\frac{z-x}{h}-\frac{y-x}{h}\right)\eta(y;\xi_2')\,\pi^2(y)\\
&\times\eta'''(z;\xi_1)\,\eta'''(z;\xi_1')\,\sigma^6(z)\,\pi(z)\,dxdydz\\
&=\frac{\Delta^3h^2}{6}\int\int\int K(p)\,\eta(x;\xi_2)\,K(p-q)\,\eta(qh+x;\xi_2')\\
&\times\eta'''(ph+x;\xi_1)\,\eta'''(ph+x;\xi_1')\,\sigma^6(ph+x)\,\pi^2(x)\,\pi^2(qh+x)\,\pi(ph+x)\,dxdpdq\\
&=\frac{\Delta^3h^2}{6}\int\int K(p)\,K(p-q)\,dpdq\\
&\times\int\eta(x;\xi_2)\,\eta(x;\xi_2')\,\eta'''(x;\xi_1)\,\eta'''(x;\xi_1')\,\sigma^6(x)\,\pi^5(x)\,dx\times[1+O(h)]
\end{aligned}$$

where the last equality holds since each $g(x):=\eta'''(x;\xi_1)\,\eta'''(x;\xi_1')\,\sigma^6(x)\,\pi(x)$ and $\pi(x)$ are globally Hölder in x (we have the Hölder coefficient uniform over ξ_1). ■