Testing Whether the Underlying Continuous-Time Process Follows a Diffusion: an Infinitesimal Operator Based Approach

Bin Chen
Department of Economics, University of Rochester

Zhaogang Song
Department of Economics, Cornell University

Preliminary Version
ABSTRACT

We develop a nonparametric test to check whether the underlying continuous time process is a diffusion, i.e., whether a process can be represented by a stochastic differential equation. Our testing procedure utilizes the infinitesimal operator based martingale characterization of diffusion models, under which the null hypothesis is equivalent to a martingale difference property of the transformed processes. Then a generalized spectral derivative test is applied to check the martingale property, where the drift function is estimated via local polynomial regression and the diffusion function is integrated out by quadratic variation. Such a testing procedure is feasible and convenient because the infinitesimal operator of the diffusion process, unlike the transition density, has a closed-form expression of the drift and diffusion functions. The proposed test is applicable to both univariate and multivariate continuous time processes and has a $N(0,1)$ limit distribution. Simulation studies show that the proposed test has good size and all-around power against non-diffusion alternatives in finite samples. We apply the test to a number of financial time series and find some evidence against the diffusion hypothesis.

Key Words: Diffusion; Generalized Spectrum; Infinitesimal Operator; Martingale; Nonparametric; Semi-group.

JEL NO: C12, C14, C22
1 Introduction

Diffusion models characterized by stochastic differential equations (SDEs) have been popularly used in modeling the dynamics of interest rates, stock prices, exchange rates and option prices over the past three decades. Continuous information flows into financial markets provide a justification for using diffusion models, and the development of stochastic calculus provides a powerful tool for elegant mathematical treatment of diffusion models. Therefore, a great number of parametric, semiparametric and nonparametric diffusion models and estimation methods have been studied in the literature (see e.g. Vasicek 1977, Cox, Ingersoll and Ross 1985, Chan, Karolyi, Longstaff and Sanders 1992, Ait-Sahalia 1996a,b, 2002, Jiang and Knight 1997, Stanton 1997, Ahn and Gao 1999, Bandi and Phillips 2003, Kristensen 2008).

While economic theories have implications on the relationship between economic variables, they usually do not suggest any concrete functional form for the processes; the choice of a model is somewhat arbitrary. Therefore, model misspecification may yield misleading conclusions on the dynamics of the process by rendering inconsistent parameter estimators and their variance-covariance matrix estimators and result in large errors in pricing, hedging and risk management in practice. To tackle such problems, many researchers have developed specification tests for diffusion models in recent years. These tests focus on the specification of the drift and diffusion either separately (see Corradi and White 1999, Li 2007 and Kristensen 2008), or jointly (e.g., Ait-Sahalia 1996a, Hong and Li 2005, Song 2008, Ait-Sahalia, Fan and Peng 2009, Chen and Hong 2010).

Although these studies can provide some guidance on choosing different specifications of the diffusion model, an implicit but essential assumption is that the underlying process is truly a diffusion, i.e., it can be represented by a SDE. Whether this assumption is supported by financial data is an interesting question. On the other hand, many empirical studies have found evidences against various parametric diffusion models (e.g., Ait-Sahalia 1996, Hong and Li 2005 and Ait-Sahalia, Fan and Peng 2009). Two possible reasons may lead to the rejection: (i) the parametric form is misspecified; (ii) the "generic diffusion hypothesis" does not hold, i.e., the underlying process itself is not a diffusion process. If the rejection is due to reason (i), we can continue searching in the class of diffusion processes for alternative parametric specification that can represent the true process. But if the rejection is from reason (ii), we should resort to other classes of processes (e.g., jump-diffusion process, Levy process and so on) instead of staying in the diffusion world and trying different parametric forms. The tests for parametric diffusion cannot disentangle cases (i) and (ii). Therefore, it is important to consider a test for the "generic diffusion hypothesis", only assuming that the underlying process is a continuous-time Markov process. Such a test can provide reliable and direct guidance on choosing suitable classes of processes in financial modeling. It can be combined with the specification tests for parametric diffusions in facilitating the modeling of financial processes: first, the test can be applied to determine whether the underlying process is a diffusion; if it is, we can try different parametric forms until a parametric form passes the specification tests; otherwise, we can try alternative classes of processes, e.g., jump-diffusion processes (Merton 1976 and Johannes 2004), Levy-driven processes, or pure jump models (Cox and Ross 1976 and Barndorff-Nielsen and Shephard 2001).

Despite the importance of testing whether the underlying process is a diffusion, only two papers consider
this issue to the best of our knowledge. Ait-Sahalia(2002) derives a restriction on the transition density of the process through the total positivity of order two property for diffusions (Karlin and McGregor 1959b). Although the total positivity condition is expressed directly for transition density and local polynomial estimators can be used to construct a test as suggested by Ait-Sahalia(2002), the partial derivatives of logarithm function complicates the problem greatly and a formal test has not been proposed yet. Ait-Sahalia(2002) only focuses on deriving such a restriction and considers some specific examples. Kanaya (2007) compares two estimators of infinitesimal operator evaluated at some specially chosen test functions, where one estimator is consistent for general Markov processes and the other is consistent only for diffusions. The test is omnibus, but an identification theorem to reduce the test function space based on the concept of a core and "approximation" theory has to be invoked. And the test statistic has an asymptotic distribution depending upon a large number of unknown parameters and functions. Moreover, it is difficult to apply the test to multivariate processes. As Kanaya (2007) comments, "it may not necessarily be an easy task as hinted by Rogers and Williams (1994, p.243): 'The moral is that for dimension $n \geq 2$, infinitesimal operators are not really the right things to look at'"

In this paper, we provide an infinitesimal operator based martingale characterization for the diffusion process and use it to construct a nonparametric test for the "generic diffusion hypothesis". The basic idea of this characterization is that when and only when an underlying process is a diffusion, a transformed process associated with the infinitesimal operator is a martingale difference sequence (MDS). This characterization is derived from the celebrated "martingale problems" in probability theory, which is developed by Strook and Varadhan(1969) to provide a "weak solution" to a SDE. As the infinitesimal operator has the closed-form expression of the drift and diffusion functions, we use nonparametric regression to estimate the drift function and the diffusion function is integrated out by the quadratic variation, which is estimated by the realized volatility. We then use a spectral approach to check whether the transformed process is a MDS. Our approach has several attractive features:

First, our test procedure is based on the convenient infinitesimal operator which is a full characterization of Markov processes and always has a closed-form expression for diffusion processes. Using the infinitesimal operator instead of the transition density not only entails no information loss for the dynamic structure of the process but also greatly simplifies the testing problem.

Second, we use a novel generalized cross-spectral derivative approach, which enjoys the appealing features of spectral analysis. In particular, our approach can examine a growing number of lags as the sample size increases without suffering from the notorious "curse of dimensionality" problem.

Third, our test is generally applicable to univariate and multivariate processes. And unlike existing tests in the literature, our test has a convenient null asymptotic $N(0,1)$ distribution. The estimation uncertainty associated with the drift and diffusion functions does not affect the asymptotic distribution of our test statistic.

There are many tests available for other generic properties of the continuous time processes, for example, the tests for jumps (e.g., Barndorff-Nielsen and Shephard 2006, Peters and de Vilder 2006, Andersen, Bollerslev, and Dobrev 2007, Ait-Sahalia and Jacod 2007) and even the tests of general Markov property (Chen and Hong 2010, Ait-Sahalia, Fan and Jiang 2010). Our test is closely related to the literature of testing for jumps: our
test can be viewed as a test for the presence of jumps because the diffusion process is essentially the only Markov process with continuous paths (see Ait-Sahalia 2002 and Kanaya 2007 for more discussion). On the other hand, other jump tests can be interpreted as a conservative test for the "generic diffusion hypothesis". If the no-jump hypothesis is rejected, then the "generic diffusion hypothesis" is also rejected. However, if not, we cannot conclude directly that the process follows a diffusion. Additional evidence is needed to support such a conclusion. Our proposed test can be used to address this more directly. Similarly, since diffusion processes belong to the category of Markov processes, the rejection of the Markov property implies the rejection of the "generic diffusion hypothesis", but not vice versa.

The paper is organized as follows. Section 2 introduces the infinitesimal operator based martingale characterization and hypotheses of interest. In Section 3, methods to smooth out the diffusion function and to estimate the drift function nonparametrically are discussed and the test is constructed by a multivariate generalized spectral derivative approach. In Section 4, we derive the asymptotic distribution of the proposed test statistic and discuss its asymptotic power property. In Section 5, we examine the finite sample performance of the test. We also apply our test to several financial time series and document some evidence against the diffusion hypothesis. Section 6 concludes. All the mathematical proofs are in the Appendix. Throughout, we use $C$ to denote a generic bounded constant, $||| \|$ for the Euclidean norm, and $A^*$ for the complex conjugate of $A$.

2 Infinitesimal Operator based Martingale Characterization

Let $\{X_t\}_{t \geq 0}$ be a $\mathbb{R}^d$-valued, time-homogeneous and continuous time Markov process defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$. What we are interested in is whether the discrete sampled data $\{X_{\tau \Delta}\}_{\tau = 1}^n$ is from a diffusion process which can be written as:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad (2.1)$$

where $W_t$ is a $d \times 1$ standard Brownian motion, $b : E \subset \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a drift function (i.e., instantaneous conditional mean), $\sigma : E \rightarrow \mathbb{R}^{d \times d}$ is a diffusion function (i.e., the instantaneous conditional standard deviation) and $E$ is the state space. We let $\mathcal{B}(E)$ be the Borel field such that $(E, \mathcal{B}(E))$ is a measurable space. The hypotheses we shall test are:

$$H_0 : \{X_t\}_{t \geq 0} \text{ is a solution to the SDE in (2.1)}$$

versus

$$H_A : \{X_t\}_{t \geq 0} \text{ is a not solution to the SDE in (2.1)}. \quad (2.2)$$

Since $\{X_t\}$ is a continuous time Markov process, we can define its transition function as $P(t, x, \Gamma) \equiv P(X_t \in \Gamma | X_0 = x)$, which is the probability that $X_t$ remains in the set $\Gamma$ starting from $x$. The Markov property is characterized by the Chapman-Kolmogorov equation: for $s, t \geq 0$, $x \in E$ and $\Gamma \in \mathcal{E}$, $P_{t+s}(x, \Gamma) = \int_E P_s(x, dy) P_t(y, \Gamma)$. An alternative and equivalent characterization is the induced family $\{P_t\}$ which is a set
of positive bounded operators with norm less than or equal to 1 on \( b(\mathcal{B}(E)) \) (bounded and \( \mathcal{B}(E) \)-measurable functions) and defined as:

\[
P_t f(x) \equiv (P_t f)(x) = \int_E P_t(x, dy) f(y).
\]  

(2.3)

In this case, the Markov property is expressed as the following semi-group property, i.e., \( P_s P_t = P_{s+t} \), for any \( s, t \geq 0 \), equivalent to the Chapman-Kolmogrov equation above. Both transition function and the semi-group of operators characterize the Markov process and interact with the sample-path property of the process. This interaction can be used to define the Feller process which contains the diffusion process in (2.1) as a special case. Let \( C_0 = C_0(E) \) defined as the space of real-valued, continuous functions on \( E \) which vanish at infinity, i.e., \( \lim_{|x| \to \infty} f(x) = 0 \), equipped with the sup-norm \( \| f \| = \sup_{x \in E} f(x) \). By Rogers and Williams (2000, Ch III.6), a process \( \{X_t\} \) is a Feller process if its semi-group of operators \( \{P_t\}_{t \geq 0} \) satisfies the following two properties: (i) \( P_tC_0 \subset C_0 \) for all \( t \geq 0 \); (ii) for any \( f \in C_0 \) and \( x \in E \), \( P_tf(x) \to f(x) \) as \( t \downarrow 0 \). Feller process has good path properties and is general enough to contain most processes we are interested in, for example, diffusion processes which have been extensively used in finance and Levy processes including Poisson process and Compound Poisson process which have received more and more attention in finance recently (see, e.g., Schoutens 2003).

For Feller processes, another characterization, namely, the infinitesimal operator, is used more frequently in probability theory than the transition function and semi-group of operators introduced above. It is defined as follows: A function \( f \in C_0 \) is said to belong to the domain \( D(\mathcal{A}) \) of the infinitesimal operator \( \mathcal{A} \) of a Feller process \( X \) if the limit

\[
\mathcal{A}f = \lim_{t \downarrow 0} \frac{P_tf - f}{t}
\]

(2.4)

exists where \( D(\mathcal{A}) \) denotes the domain of \( \mathcal{A} \), i.e., the family of functions in \( C_0 \) for which the limit in (2.4) exists with respect to the sup-norm of \( C_0 \). Obviously, \( \mathcal{A} \) is a linear operator from \( D(\mathcal{A}) \) to \( C_0 \). It can be seen from (2.4) immediately, that it holds \( P\text{-a.s.} \) for \( f \in D(\mathcal{A}) \)

\[
E\left( \frac{f(X_{t+\Delta}) - f(X_t)}{\Delta} \bigg| \mathcal{F}_t \right) = \mathcal{A}f(X_t) + o(\Delta), \text{ as } \Delta \to 0.
\]

(2.5)

In this sense, the infinitesimal operator indeed describes the movement of the process in an infinitesimally small amount of time. Therefore, the infinitesimal operator characterizes the whole dynamics of a Feller process because the time is continuous. It can be proved that the infinitesimal operator is equivalent to the semi-group of operators in characterizing a Feller process (see the Hill-Yoshida theorem in Dynkin 1965). By the equivalence of semi-group of operators and the transition function, the infinitesimal operator determines the transition function and thus fully characterizes the dynamics of the process.

---

1. By Rogers and Williams (2000, Ch III.7-9), the canonical Feller process always admits a Cadlag (the path of the process is right continuous and has left limits) modification and satisfies the strong Markov property.
2. As pointed out by Kanaya (2007), there exists another way to define the infinitesimal operator without using the sup-norm. For example, Hansen and Scheinkman (1995) define infinitesimal operator in the Hilber space \( L^2(Q) \) where \( Q \) is an invariant(stationary) distribution of the process. This Hilber space based definition is needed in Hansen and Scheinkman (1995) for analyzing such properties as time reversibility. But unlike their method, our approach does not need the assumption of time reversibility. Therefore, the definition using \( C_0 \) is enough and our method is less restricted.
For the diffusion process in (2.1), the infinitesimal operator enjoys the nice property of being always a closed-form expression which can be identified by the drift and diffusion functions. According to Kallenberg (2002, Thm 19.24) and Rogers and Williams (2000, Vol1, Thm III.13.3 and Vol2, Ch V.2), the infinitesimal operator of the diffusion model in (2.1) is:

\[
\mathcal{A}_\theta f(x) = \sum_{i=1}^{d} b_i(x; \theta) f'_i(x) + \frac{1}{2} \sum_{i,j=1}^{d} a_{i,j}(x) f''_{i,j}(x), \ f \in D(\mathcal{A}), \ x \in \mathbb{R}^d
\]

(2.6)

where

\[
a_{ij}(x) = \sum_{k=1}^{d} \sigma_{i,k}(x) \sigma_{j,k}(x).
\]

For illustration, we consider a univariate diffusion model \(dX_t = b(X_t)dt + \sigma(X_t)dW_t\), where \(W_t\) is a one-dimensional standard Brownian motion in \(\mathbb{R}\), \(b : E \subset \mathbb{R} \to \mathbb{R}\) is a drift function and \(\sigma : E \to \mathbb{R}\) is a diffusion function. Then by (2.6) and the discussions above, the infinitesimal operator for this univariate diffusion is

\[
\mathcal{A}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x).
\]

(2.7)

Clearly the first and second terms are related to the dynamics of drift and diffusion functions respectively. This is consistent with the intuition that the drift describes the dynamics of mean and the diffusion describes that of the variance of the process (see Nelson 1990 for more discussion). Of course, this intuition should not be taken literally due to the continuous nature of the time. Consider the infinitesimal changes of this univariate diffusion process. By (2.5) and (2.7), for any \(f \in D(\mathcal{A})\), it holds \(P\)-a.s. that

\[
E \left( \frac{f(X_{t+\Delta}) - f(X_t)}{\Delta} | \mathcal{F}_t \right) = b(X_t)f'(X_t) + \frac{1}{2}\sigma^2(X_t)f''(X_t) + o(\Delta), \text{ as } \Delta \to 0.
\]

Therefore, the dynamics of \(\{X_t\}\) are characterized completely by the drift and diffusion coefficients, including the conditional probability law. In contrast, for discrete time models, the mean and variance cannot solely determine the conditional probability law unless it is a Gaussian process. Therefore, it is incorrect to simply view the drift and diffusion as the straightforward continuous time counterparts of the conditional mean and variance respectively. In fact, the conditional mean of the process \(\{X_t\}\), \(E[X_{t+\Delta}|X_t]\) for a fixed \(\Delta > 0\) is a function of both the drift \(b(\cdot)\) and diffusion \(\sigma(\cdot)\) instead of the drift solely (see Ait-Sahalia 1996a). The precise expressions for drift and diffusion functions are:

\[
b(X_t) = \lim_{\Delta \to 0} E \left[ \frac{X_{t+\Delta} - X_t}{\Delta} | X_t \right],
\]

\[
\sigma^2(X_t) = \lim_{\Delta \to 0} E \left[ \frac{(X_{t+\Delta} - X_t)^2}{\Delta} | X_t \right],
\]

(2.8)

which are called instantaneous conditional mean and variance\(^3\).

\(^3\)These definitions are employed by Stanton(1997) and Bandi and Phillips(2003) to propose nonparametric estimators for drift and diffusion functions.
Since the diffusion process in (2.1) is a Feller process, we have three complete characterizations of the dynamics available: transition function (or transition density), semi-group of operators and infinitesimal operator. The transition function has already been used intensively in econometric inference of diffusion models, not only in estimation (Lo 1988; Ait-Sahalia 2002; Pedersen 1995) but also in hypothesis testing (Ait-Sahalia, Fan and Peng 2008, Hong and Li 2005;). However, as we know, the transition density of most continuous time models has no closed form. Therefore, those methods based on the transition density are usually computationally burdensome and inconvenient to be applied in practice. In contrast, the infinitesimal operator of a diffusion process always has a closed-form and fully characterizes the dynamics. This nice property, therefore, makes the infinitesimal operator a convenient tool for analyzing the diffusion models. It has already been employed for identification (Hansen, Scheinkman and Touzi 1998), estimation (Hansen and Scheinkman 1995; Kessler and Sorenson 1999, Song 2009) and also specification testing (Kanaya 2007, Song 2008).

To obtain a convenient characterization which can be used to construct a test, we consider a transformation based on the celebrated "martingale problems". This transformation gives us a martingale characterization for diffusion processes which is not only a complete identification but also very simple and convenient to use. By Ch5.4 of Karatzas and Shreve (1991), a probability measure $P$ on $(C[0,\infty)^d, \mathcal{B}(C[0,\infty)^d))$ under which

$$M^f_t = f(X_t) - f(X_0) - \int_0^t (Af)(X_s) ds$$

is a martingale for every $f \in D(A)$

(2.9)

is called a solution to the martingale problem associated with the operator $A$. As we know, a SDE has two types of solutions: strong solutions and weak solutions (see Karatzas and Shreve 1991, Ch5.2-3 or Rogers and Williams 2000, ChV.2-3 for details). Intuitively, the strong solution is a solution to SDE with a.s.properties and a weak solution is that to SDE with in law properties. When the drift and diffusion terms of a SDE satisfy the Lipschitz and linear growth conditions, there is a strong solution to the SDE. But in the general case, a strong solution may not exist, where probabilists usually attempt to solve the SDE in the "weak" sense of finding a solution with the right probability law. The martingale problem is a variation of this "weak solution approach" developed by Strook and Varadhan (1969) and is equivalent to the weak solution of a SDE. For detailed discussions and proof, see ChV.19-20 of Rogers and Williams (2000), or Theorem 21.7 of Kallenberg (2002), or Proposition 2.4 of ChVII in Revuz and Yor (2005). One thing we should point out is that when strong solution exists the weak solution will coincide with it. Hence it is enough to consider the weak solution identification for econometric inferences because regularity conditions for the existence of strong solution are usually satisfied and thus imposed in analysis\(^4\). See Protter (2005) for some regularity Lipschitz conditions for the existence and uniqueness of a strong solution to a SDE. By (2.9), the hypotheses in (2.2) can be equivalently written as:

$$H_0 : M^f_t = f(X_t) - f(X_0) - \int_0^t (Af)(X_s) ds$$

is a martingale for every $f \in D(A)$

\(^4\)We thank Philip Protter for suggesting this point.
versus
\[ \mathbb{H}_A : M^f_t = f(X_t) - f(X_0) - \int_0^t (Af)(X_s)ds \] is not a martingale for some \( f \in D(A) \),

where
\[ Af(x) = \sum_{i=1}^d b_i(x)f'_i(x) + \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x)f''_{ij}(x) \text{ and } a_{ij}(x) = \sum_{k=1}^d \sigma_{i,k}(x)\sigma_{j,k}(x). \] (2.10)

Therefore, the identification of the diffusion process in (2.1) is equivalent to the martingale property of the transformed processes in (2.9). However, observe that we have infinite many transformed processes \( \{M^f_t\}_{f \in D(A)} \) because there are usually an infinite number of functions \( f(\cdot) \) in the domain \( D(A) \) which are called test functions. It is very difficult and burdensome, although maybe not impossible, to check simultaneously that they are all martingales. This is a general problem which appears not only in our study here but also for all the other papers employing infinitesimal operators, like Hansen and Scheinkman (1995), Conley, Hansen, Luttmer and Scheinkman (1997), and Kanaya (2007). To tackle such a difficulty, the space of test functions has to be reduced to an equivalent subclass. Kanaya (2007) does this based on the concept of a core and "approximation" theory. Hansen and Scheinkman (1995) and Conley, Hansen, Luttmer and Scheinkman (1997) also discuss the choice of test functions. But no formal evidence is provided for the equivalence and identification.

In contrast, we shall depend on a celebrated theorem in the probability theory to obtain a subclass of \( D(A) \) which not only consists of finitely many function forms but also plays the same role as \( D(A) \) for the martingale characterization (2.9). By Proposition 4.6 and Remark 4.12 of Karatzas and Shreve (1991, Ch5.4), the process \( \{X_t\} \) is a weak solution to the SDE in (2.1) if it satisfies the martingale problem with \( A \) as the infinitesimal operator of \( \{X_t\} \) for the choices \( f(x) = x_i \) and \( f(x) = x_i x_j \) with \( 1 \leq i, j \leq d \). According to this, the hypotheses of interest in (2.10) can be transformed as:

\[ \mathbb{H}_0 : \text{The transformed processes} \]

\[ M^x_t = X^i_t - X^0_i - \int_0^t b_i(X_s)ds \]

\[ M^{x_i,x_i}_t = (X^i_t)^2 - (X^0_i)^2 - \int_0^t \left[ 2b_i(X_s)X^i_s + \sum_{k=1}^d \sigma_{i,k}(X_s)^2 \right] ds \]

\[ M^{x_i,x_j}_t = X^i_t X^j_t - X^0_i X^0_j - \int_0^t \left[ b_i(X_s)X^j_s + b_j(X_s)X^i_s + \frac{1}{2} \sum_{k=1}^d \sigma_{i,k}(X_s)\sigma_{j,k}(X_s) \right] ds \text{ for } i \neq j \] (2.11)

are martingales for \( 1 \leq i, j \leq d \).

Versus

\[ \mathbb{H}_A : \text{either } M^x_t \text{ or } M^{x_i,x_j}_t \text{ is not a martingale for some } i, j = 1, \ldots, d. \]

The characterization (2.11) is much simpler and more intuitive than other characterization in the literature (e.g., Hansen and Scheinkman 1995, Kanaya 2007). It greatly simplifies the hypotheses of interest and makes the test for the generic diffusion hypothesis feasible. The hypotheses can be expressed explicitly by the drift and diffusion terms. Therefore, they can be used directly while in contrast, the transition density based
methods (e.g., Ait-Sahalia 2002) need to either approximate the transition density or numerically solve it because the transition density rarely has a closed-form. On the other hand, the identification of a multivariate \(d\)-dimensional diffusion process is equivalent to the martingale property for \(d'(d^2 + 3d)/2\) univariate processes which are explicit expressions of drift and diffusion terms. This makes the transformed equivalent hypotheses particularly convenient for multivariate diffusion models for which the transition density methods are extremely complicated and computationally inconvenient.

For the convenience of constructing a test procedure, we further state the following equivalent hypotheses in terms of the MDS property for the transformed processes

\[
H_0 : E[Z_t | \mathcal{I}_t] = 0 \text{ for any } t' < t, \tag{2.12}
\]

where \(\mathcal{I}_t = \sigma \{X_{t'} \}_{t' < t} \) is the sigma-field generated by the past information of \(\{X_t\}\) at time \(t'\) and \(Z_t\) is a vector with the following components, for \(i, j = 1, \cdots, d\),

\[
Z_t^i = M_t^{X^i} - M_{t-\Delta}^{X^i} = X_t^i - X_{t-\Delta}^i - \int_{t-\Delta}^{t} b_i(X_s) ds
\]

\[
Z_t^{i,j} = M_t^{X^{iX^j}} - M_{t-\Delta}^{X^{iX^j}} = (X_t^i)^2 - (X_{t-\Delta}^i)^2 - \int_{t-\Delta}^{t} \left[ 2b_i(X_s)X_s^i + \sum_{k=1}^{d} \sigma_{i,k}(X_s)^2 \right] ds + \int_{t}^{t-\Delta} \left[ b_i(X_s)X_s^i + b_j(X_s)X_s^j + \frac{1}{2} \sum_{k=1}^{d} \sigma_{i,k}(X_s)\sigma_{j,k}(X_s) \right] ds \quad \text{for } i \neq j \tag{2.13}
\]

versus

\[H_A : E[Z_t | \mathcal{I}_t] \neq 0 \text{ for some } t' < t. \tag{2.14}\]

We note that as a special case, the MDS representation of the diffusion hypothesis for univariate case is:

\[H_0 : E[Z_t | \mathcal{I}_t] = 0 \text{ for any } t' < t, \text{ where } Z_t = \begin{pmatrix} Z_t^X \\ Z_t^{X^2} \end{pmatrix}, \mathcal{I}_t = \sigma \{X_{t'} \}_{t' < t}, \text{ and} \]

\[
Z_t^X = M_t^X - M_{t-\Delta}^X = X_t - X_{t-\Delta} - \int_{t-\Delta}^{t} b(X_s) ds
\]

\[
Z_t^{X^2} = M_t^{X^2} - M_{t-\Delta}^{X^2} = X_t^2 - X_{t-\Delta}^2 - \int_{t-\Delta}^{t} \left[ 2b(X_s)X_s + \sigma^2(X_s) \right] ds \tag{2.15}
\]

versus

\[H_A : E[Z_t(\theta) | \mathcal{I}_t] \neq 0 \text{ for some } t' < t. \tag{2.16}\]

### 3 Testing Procedure

In this section, we shall construct the test procedure of the hypotheses \(H_0\) versus \(H_A\) in (2.13) and (2.14) for the multivariate continuous time processes. The sample data is discrete in time, i.e., \(\{X_t \}_{t=1}^{n}\) observed over a time span \(T\) with sampling interval \(\Delta\) and sample size \(n = T/\Delta\). This is a general problem in continuous-time series econometrics (see Lo 1988 and Ait-Sahalia 1996 for discussions about the estimation of the discretized
continuous-time models). The asymptotic schemes we employ are $n = T/\Delta \to \infty$ and for each sampling interval $\Delta$, we have high frequency data with the sampling interval $\delta = \Delta/M \to 0$ for integer $M$. The former is a standard treatment in the literature of estimating and testing diffusion models (Ait-Sahalia, 1996a,b; Ait-Sahalia, Fan, and Peng, 2009; Hong and Li, 2005) while the latter is to ensure the consistency of realized volatility (covariation) to the quadratic variation (covariation) where we integrate out the unknown diffusion function.

The null hypothesis in (2.13) is a conditional moment restriction, i.e., $E[Z_t|I_{t'}] = 0$ for any $t' < t$, where $I_{t'} = \sigma\{X_{t''}\}_{t'' < t'}$ is the sigma-field generated by the past information of $\{X_t\}$ and $Z_t$ is a vector with components defined in (2.13). By the relationship between $Z_t$ and $X_t$ in (2.13), we have

$$E[Z_t|I_{t'}] = 0 \text{ for any } t' < t, \text{ for } I_{t'} = \sigma\{Z_{t''}\}_{t'' < t'},$$

where $I_{t'}$ is the sigma-field generated by past information of $\{Z_t\}$. Since the sample data we have is $\{X_{\tau\Delta}\}_{\tau = 1}$ with $n = T/\Delta$, an application of the law of iterated expectation as well as (3.1) implies that

$$E[Z_{\tau\Delta}|I_{(\tau-1)\Delta}] = 0, \text{ where } I_{(\tau-1)\Delta} = \sigma\{Z_{(\tau-1)\Delta}, Z_{(\tau-2)\Delta}, \cdots, Z_{\Delta}\}.$$  

We note that (3.2) is a MDS property for discrete time process $\{Z_{\tau\Delta}\}_{\tau = 1}$ and it is derived as an implication of the MDS property in continuous time instead of a result from the discretization of the continuous time process. In this respect, it is similar to the approaches of Ait-Sahalia (1996a,b) and Lo (1988) and free of the discretization errors (see Lo 1988 for detailed discussion).

Our test procedure is based on the MDS characterization in Eq. (3.2). However, it is not a trivial task to check it. First, the conditioning information set $I_{\tau-1\Delta}$ has an infinite dimension as $\tau \to \infty$ and then there is a "curse of dimensionality" difficulty associated with testing the MDS property. Second, $\{Z_{\tau\Delta}\}$ may display serial dependence in its higher order conditional moments. Any test should be robust to time-varying conditional heteroskedasticity and higher order moments of unknown form in $\{Z_{\tau\Delta}\}$. To check the MDS property of $\{Z_{\tau\Delta}\}$, we apply a multivariate generalized spectral derivative approach, which is an extension of the generalized spectral method in Hong (1999) and Hong and Lee (2005).

Suppose $\{Z_{\tau\Delta}\}$ is a strictly stationary process with marginal characteristic function $\varphi(u) = E(e^{iu'Z_{\tau\Delta}})$ and pairwise joint characteristic function $\varphi_m(u, v) = E(e^{iu'Z_{\tau\Delta}+iv'Z_{(\tau-|m|)\Delta}})$, where $i = \sqrt{-1}$, $u, v \in \mathbb{R}^d$, and $m = 0, \pm 1, \cdots$. The basic idea of the generalized spectrum is to consider the spectrum of the transformed series $\{e^{iu'Z_{\tau\Delta}}\}$. It is defined as

$$f(\omega, u, v) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sigma_m(u, v)e^{-im\omega}, \omega \in [-\pi, \pi]$$
where $\omega$ is the frequency, and $\sigma_m(u, v)$ is the covariance function of the transformed series:

$$\sigma_m(u, v) \equiv \text{cov}(e^{iu'Z_{r\Delta}} , e^{iv'Z_{(r-m)\Delta}}), m = 0, \pm 1, \cdots. \quad (3.3)$$

Note that the function $f(\omega, u, v)$ is a complex-valued scalar function although $Z_{r\Delta}$ is a $d' \times 1$ vector. It can capture any type of pairwise serial dependence in $\{Z_{r\Delta}\}$, i.e., dependence between $Z_{r\Delta}$ and $Z_{(r-m)\Delta}$ for any nonzero lag $m$, including that with zero autocorrelation. First, this is analogous to the higher order spectra (Brillinger and Rosenblatt, 1967a,b) in the sense that $f(\omega, u, v)$ can capture the serial dependence in higher order moments. However, unlike the higher order spectra, $f(\omega, u, v)$ does not require the existence of any moment of $\{Z_{r\Delta}\}$. Second, this can capture nonlinear dynamics while maintaining the nice features of spectral analysis, especially its appealing property to accommodate information in all lags. In the present context, it can check the MDS property over many lags in a pairwise manner, avoiding the "curse of dimensionality" problem. This is not achievable by other existing tests in the literature which only check a fixed lag order.

The generalized spectrum $f(\omega, u, v)$ itself cannot be applied directly for testing $H_0$, because it captures the serial dependence not only in mean but also in higher order moments. However, just as the characteristic function can be differentiated to generate various moments of the serial dependence not only in mean but also in higher order moments. However, just as the characteristic function can be differentiated to capture any type of pairwise serial dependence in $\{Z_{r\Delta}\}$, $f(\omega, u, v)$ can be differentiated to capture the serial dependence in various moments. To capture (and only capture) the serial dependence in conditional mean, one can consider the derivative:

$$f^{(0,1,0)}(\omega, 0, v) \equiv \frac{\partial}{\partial u} f(\omega, u, v)|_{u=0} = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sigma_m^{(1,0)}(0, v)e^{-im\omega}, \omega \in [-\pi, \pi]$$

where

$$\sigma_m^{(1,0)}(0, v) \equiv \frac{\partial}{\partial u}\sigma_m(u, v) \mid_{u=0} = \text{cov}(iZ_{r\Delta}, e^{iv'Z_{(r-m)\Delta}}) \quad (3.4)$$

is a $d' \times 1$ vector. The measure $\sigma_m^{(1,0)}(0, v)$ checks whether the autoregression function $E[Z_{r\Delta} \mid Z_{(r-m)\Delta}]$ at lag order $m$ is zero. Under some regularity conditions, $\sigma_m^{(1,0)}(0, v) = 0$ for all $v \in \mathbb{R}^{d'}$ if and only if $E[Z_{r\Delta} \mid Z_{(r-m)\Delta}] = 0$, a.s..

It should be noted that the hypothesis of $E \left[ Z_{r\Delta} \mid T_{(r-1)\Delta}^{2} \right] = 0$ a.s. is not exactly the same as the hypothesis of $E[Z_{r\Delta} \mid Z_{(r-m)\Delta}] = 0$ a.s. for all $m \neq 0$. The former implies the latter but not vice versa. There exists a gap between them. This is the price we have to pay to deal with the difficulty of the "curse of dimensionality". Nevertheless, the examples where $E[Z_{r\Delta} \mid Z_{(r-m)\Delta}] = 0$ a.s. for all $m \neq 0$ but $E \left[ Z_{r\Delta} \mid T_{(r-1)\Delta}^{2} \right] \neq 0$ a.s. may be rare in practice and are thus pathological. Even for cases where the gap does matter, it can be further narrowed down by using the function $E[Z_{r\Delta} \mid Z_{(r-m)\Delta}, Z_{(r-l)\Delta}]$ which may be called the bi-autoregression function of $Z_{r\Delta}$ at lags $(m, l)$. An equivalent measure is the generalized third order central cumulant function $\sigma_m^{(1,1)}(0, v) = \text{cov} \left[ Z_{r\Delta}, \exp \left( iv_1Z_{(r-m)\Delta} + iv_2Z_{(r-l)\Delta} \right) \right]$, where $v = (v_1, v_2) \in \mathbb{R}^{d'} \times \mathbb{R}^{d'}$. This is a straightforward extension of generalized bispectrum analysis proposed by Hong and Song (2008) for univariate processes.

As the transformed process $\{Z_{r\Delta}\}$ is not observable, we have to estimate it. We do not parameterize $b(\cdot)$ or $\sigma(\cdot)$, which would suffer from potential model misspecification. We shall use nonparametric regression to estimate them. First we consider local polynomial regression to estimate the drift function as it can be viewed
as the instantaneous conditional mean; namely
\[
b(x) = \lim_{\Delta \to 0} E \left[ \frac{X_{(r+1)\Delta} - X_{r\Delta}}{\Delta} \middle| X_{r\Delta} = x \right].
\]

Local polynomial smoothing is introduced originally by Stone (1977) and subsequently studied by Cleveland (1979), Fan (1992, 1993), Ruppert and Wand (1994), Masry (1996a, 1996b) and Masry and Fan (1997), among many others. Local polynomial smoothing has some advantages over the conventional Nadaraya–Watson (NW) kernel estimator: e.g., local polynomial fits adapt automatically to the boundary regions when the order of polynomial \( r \) is odd (Ruppert and Wand 1994, Fan and Yao 2003); it is superior to the NW estimator in the context of estimating the derivatives of the regression function (Ruppert and Wand 1994, Fan and Yao 2003).

Following Masry (1996a, 1996b), we introduce the following notations:
\[
j = (j_1, \ldots, j_d), \quad j! = j_1! \times \cdots \times j_d!, \quad |j| = \sum_{i=1}^d j_i,
\]
\[
x^{j} = x^{j_1}_1 \times \cdots \times x^{j_d}_d,
\]
\[
\sum_{0 \leq |j| \leq r} = \sum_{l=0}^{r} \sum_{j_1=0}^{l} \cdots \sum_{j_d=0}^{l}.
\]

We consider the following multivariate local weighted least squares problem:
\[
\min_{\beta \in \mathbb{R}^N} \sum_{\tau=2}^n \left| \frac{X_{\tau\Delta} - X_{(\tau-1)\Delta}}{\Delta} - \sum_{0 \leq |j| \leq r} \beta_j^\tau (X_{(\tau-1)\Delta} - x)^j \right|^2 K_h [X_{(\tau-1)\Delta} - x], \quad x \in \mathbb{R}^d, \tag{3.5}
\]
where \( \beta = (\beta_0, \beta'_1, \ldots, \beta'_r)' \) is an \( N \times 1 \) parameter vector, \( N = \sum_{l=0}^r N_l, \quad N_l = \frac{(l+d-1)!}{(d-1)!l!} \), \( K_h (x) = h^{-d}K(x/h), \quad K : \mathbb{R}^d \to \mathbb{R} \) is a kernel function, \( h \) is a bandwidth and \( r \) is an odd integer. When \( r = 1 \), Eq. (3.5) boils down to a local linear regression. An example of \( K(\cdot) \) is a prespecified symmetric probability density function. We obtain the following solution to Eq. (3.5):
\[
\hat{\beta} \equiv \hat{\beta}(x, u) = \begin{bmatrix} \hat{\beta}_0 (x, u) \\ \vdots \\ \hat{\beta}_r (x, u) \end{bmatrix} = S_{T}^{-1}(x) \Gamma (x, u), \quad x \in \mathbb{R}^d,
\]
where \( S_T (x) \) is an \( N \times N \) matrix
\[
S_T (x) = \begin{bmatrix} S_{0,0} & S_{0,1} & \cdots & S_{0,r} \\ S_{1,0} & S_{1,1} & \cdots & S_{1,r} \\ \vdots & \vdots & \ddots & \vdots \\ S_{r,0} & S_{r,1} & \cdots & S_{r,r} \end{bmatrix},
\]
$S_{[j], [l]}$ is an $N_{[j]} \times N_{[l]}$ matrix with its $(m, n)$th element $(S_{[j], [l]})_{m,n} = s_{g_{[j]}(m) + g_{[l]}(n)}$.

\[
s_j(x) = \frac{1}{n-1} \sum_{\tau=2}^{n} \left( \frac{X_{(\tau-1)\Delta} - x}{h} \right)^j K_h(X_{(\tau-1)\Delta} - x)
\]

and $g^{-1}_j$ denotes the one-to-one map that arranges those $N_j$ $d$-tuples as a sequence in a lexicographical order.\(^6\)

And $\Gamma(x)$ is an $N \times 1$ vector

\[
\Gamma(x) = \begin{bmatrix} 
\Gamma_0 \\
\Gamma_1 \\
\vdots \\
\Gamma_r 
\end{bmatrix},
\]

$\Gamma_{[j]}$ is of dimension $N_{[j]} \times 1$, with its $l$–th element $(\Gamma_{[j]})_l = \tau_{g_{[j]}(l)}$, and

\[
\tau_j(x) = \sum_{\tau=2}^{n} \frac{X_{\tau\Delta} - X_{(\tau-1)\Delta}}{\Delta} \left( \frac{X_{(\tau-1)\Delta} - x}{h} \right)^j K_h(X_{(\tau-1)\Delta} - x).
\]

Note that $\hat{\beta}$ depends on the location $x$, but for notional simplicity, we have suppressed its dependence on $x$.

Under suitable regularity conditions, $b(x)$ can be consistently estimated by the local intercept estimator $\hat{\beta}_0(x)$. Specifically, we have

\[
\hat{b}(x) = \sum_{\tau=2}^{n} \hat{W} \left[ \frac{X_{(\tau-1)\Delta} - x}{h} \right] \left[ \frac{X_{\tau\Delta} - X_{(\tau-1)\Delta}}{\Delta} \right],
\]

where $\hat{W}(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is an effective kernel, defined as

\[
\hat{W}(z) \equiv (n-1)^{-1} e_1^T S_n^{-1} \Theta(z) K(z) / h^d,
\]

$e_1 = (1, 0, ..., 0)'$ is an $N \times 1$ unit vector, $\Theta(z)$ is an $N \times 1$ vector

\[
\Theta(z) = \begin{bmatrix} 
\Theta_0(z) \\
\Theta_1(z) \\
\vdots \\
\Theta_r(z) 
\end{bmatrix},
\]

$\Theta_{[j]}(z)$ is of dimension $N_{[j]} \times 1$, with its $l$–th element $(\Theta_{[j]}(z))_l = (z)^{g_{[j]}(l)}$ and $z$ is a $d \times 1$ vector. The regression estimator $\hat{b}(x)$ only involves a $d$-dimensional smoothing, thus enjoying some advantages over the existing nonparametric density approaches which involve a $2d$ or $3d$ dimensional smoothing.

For the diffusion function $\sigma^2(\cdot)$, we could apply the similar approach,

\[
\sigma^2(x) = \sum_{\tau=2}^{n} \hat{W} \left[ \frac{X_{(\tau-1)\Delta} - x}{h} \right] \left[ \frac{X_{\tau\Delta} - X_{(\tau-1)\Delta}}{\Delta} \right]^2.
\]

\(^6\)See Masry (1996a, 1996b) for the detailed explanation for these notations.
where $\hat{W}(\cdot)$ is defined in (3.7). However, the infinitesimal operator based martingale characterization employed gives us a specific alternative choice based on quadratic variation (covariation). We observe from (2.13) that the diffusion function only appears in $Z_t^{i,i}$ and $Z_t^{i,j}$ with the forms $\int_{t-\Delta}^{t} \left[ \sum_{k=1}^{d} \sigma_{i,k}^2(X_s) \right] ds$ and $\int_{t-\Delta}^{t} \left[ \sum_{k=1}^{d} \sigma_{i,k}(X_s) \sigma_{j,k}(X_s) \right] ds$ (for $i \neq j$) respectively. It is well known that for the diffusion models, the former is equal to the quadratic variation $[X_i, X_i]^{t}_{t-\Delta}$ (also known as integrated volatility) and the latter is the quadratic covariation $[X_i, X_j]^{t}_{t-\Delta}$, which have been analyzed intensively in recent years (see Andersen, Bollerslev, Diebold, and Labys, 2003, Barndorff-Nielsen and Shephard 2004, 2006, Ait-Sahalia, Mykland and Zhang 2005, Barndorff-Nielsen and Shephard 2004a, Bandi and Russell 2005 and Zhang 2006). Now change the notations and then we have

$$Z_t^{i,i} = Z_t^{i,i}([X_i, X_i]) = (X_i^i)^2 - (X_i^{i-})^2 - \int_{t-\Delta}^{t} \left[ 2b_i(X_s)X_s^i \right] ds - \left[ X_i^i, X_i^i \right]^{t}_{t-\Delta}$$

$$Z_t^{i,j} = Z_t^{i,j}([X_i, X_j]) = X_i^iX_j^j - X_i^{i-}X_j^{j-} - \int_{t-\Delta}^{t} \left[ b_i(X_s)X_s^i + b_j(X_s)X_s^j \right] ds - \frac{1}{2} \left[ X_i^i, X_j^j \right]^{t}_{t-\Delta} \text{ for } i \neq j. \quad (3.9)$$

Therefore the diffusion function is integrated out by the quadratic variation and covariation which can be estimated consistently by the realized volatility and covariance respectively:

$$[\hat{X}^i, \hat{X}^i]^{t}_{t-\Delta} = \sum_{i=1}^{M} \left( X_{t-\Delta+i, \Delta - (i-1) \delta}^i - X_{t-\Delta, \Delta - (i-1) \delta}^i \right)^2$$

$$[\hat{X}^i, \hat{X}^j]^{t}_{t-\Delta} = \sum_{i=1}^{M} \left( X_{t-\Delta+i, \Delta - (i-1) \delta}^i - X_{t-\Delta, \Delta - (i-1) \delta}^i \right) \left( X_{t-\Delta+i, \Delta - (i-1) \delta}^j - X_{t-\Delta, \Delta - (i-1) \delta}^j \right) . \quad (3.10)$$

Of course, the infill asymptotic scheme has to be assumed, i.e., $\delta \to 0$. Consequently, we have

$$[\hat{X}^i, \hat{X}^i]^{t}_{t-\Delta} - [X^i, X^i]^{t}_{t-\Delta} = O_p(\delta^{1/2}) \quad (3.11)$$

and

$$[\hat{X}^i, \hat{X}^j]^{t}_{t-\Delta} - [X^i, X^j]^{t}_{t-\Delta} = O_p(\delta^{1/2}) \quad (3.12)$$


Compared with the local polynomial estimator in (3.8), this approach of integrating out diffusion functions by quadratic variation and covariance has some advantages. First, the quadratic variation and covariation method cover the multivariate cases easily and are free of the "curse of dimensionality" which is generally suffered by nonparametric smoothing methods. Second, the realized volatility and covariance are essentially nonparametric and do not involve the choice of any other parameters. In contrast, one has to choose the kernel function and smoothing bandwidth in nonparametric smoothing methods and the choice of the latter is nontrivial. Third, the estimators for the quadratic variation and covariance, i.e., the realized volatility and covariance have faster convergence rate than the nonparametric smoothing estimator for the diffusion function.
if $M > Th^d$. This will lead to smaller estimation uncertainty and thus better finite sample performance of the test statistic. Of course, since high frequency data are needed to estimate the quadratic variation and covariation, microstructure noises could appear. In this case, we could follow Zhang, et al. (2005) to effectively combine the average of realized volatilities estimated over subgrids on a slow time scale with the realized volatility estimated using all of the data. Zhang et al. (2005) show that the bias-corrected estimator is consistent despite the presence of market microstructure noise and the convergence rate could be $(M/\Delta)^{\frac{1}{2}}$ if certain optimal number of subgrids is chosen.

With $\hat{b}(\cdot)$ and $[\hat{X}, \hat{X}]$ in hand, we still need to approximate the Lebesgue integrals in (2.11) to compute $Z_{\tau\Delta}(\cdot)$ for our test. A simple approximation is

$$
\int_{(\tau-1)\Delta}^{\tau\Delta} b_i(X_s)ds = \frac{\Delta}{2} \left\{ b_i(X_{\tau\Delta}) + b_i(X_{(\tau-1)\Delta}) \right\} + O_P(\Delta^2)
$$

and

$$
\int_{(\tau-1)\Delta}^{\tau\Delta} 2b_i(X_s)X_s^i ds = \Delta \left\{ X_{\tau\Delta}^i b_i(X_{\tau\Delta}) + X_{(\tau-1)\Delta}^i b_i(X_{(\tau-1)\Delta}) \right\} + O_P(\Delta^2).
$$

A more complicated approximation is

$$
\int_{(\tau-1)\Delta}^{\tau\Delta} b_i(X_s)ds = \frac{\Delta}{M} \sum_{m=(\tau-1)\Delta + \frac{\Delta}{M}}^{\tau\Delta} b_i(X_m) + O_P \left( M^{-\frac{1}{2}} \right)
$$

and

$$
\int_{(\tau-1)\Delta}^{\tau\Delta} 2b_i(X_s)X_s^i ds = \frac{\Delta}{M} \sum_{m=(\tau-1)\Delta + \frac{\Delta}{M}}^{\tau\Delta} 2b_i(X_m)X_m^i + O_P \left( M^{-\frac{1}{2}} \right).
$$

Similar approximations would be employed for other Lebesgue integrals in (2.11).

Given the estimated drift and quadratic variation (covariation) and numerically approximations of the Lebesgue integrals, we can obtain the estimated transformed processes $\hat{Z}_{\tau\Delta}$. Then we can estimate $f^{(0,1,0)}(\omega, 0, v)$ for process $\{Z_{\tau\Delta}\}$ by the following smoothed kernel estimator:

$$
\hat{f}^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{m=1-n}^{n-1} (1-|m|/n)^{1/2} k(m/p)\hat{\sigma}^{(1,0)}(0, v)e^{-im\omega}, \omega \in [-\pi, \pi] \text{ and } v \in \mathbb{R}^d,
$$

where $\hat{\sigma}^{(1,0)}(0, v) \equiv \frac{\partial}{\partial u} \hat{\sigma}_m(u, v) |_{u=0}$, $\hat{\sigma}_m(u, v) = \hat{\varphi}_m(u, v) - \hat{\varphi}_m(u, 0)\hat{\varphi}_m(0, v)$,

$$
\hat{\varphi}_m(u, v) = \frac{1}{n-|m|} \sum_{\tau=|m|+1}^{n} e^{iu\tau + iv\hat{Z}_{\tau\Delta} + iv\hat{Z}_{(\tau-|m|)\Delta}},
$$

where $\hat{\sigma}^{(1,0)}(0, v) \equiv \frac{\partial}{\partial u} \hat{\sigma}_m(u, v) |_{u=0}$, $\hat{\sigma}_m(u, v) = \hat{\varphi}_m(u, v) - \hat{\varphi}_m(u, 0)\hat{\varphi}_m(0, v)$,

$$
\hat{\varphi}_m(u, v) = \frac{1}{n-|m|} \sum_{\tau=|m|+1}^{n} e^{iu\tau + iv\hat{Z}_{\tau\Delta} + iv\hat{Z}_{(\tau-|m|)\Delta}},
$$

$p = p(n)$ is a bandwidth, and $k : \mathbb{R} \rightarrow [-1, 1]$ is a symmetric kernel. Examples of $k(\cdot)$ include Bartlett, Daniell, Parzen and Quadratic spectral kernels (e.g., Priestley 1981, p.442). The factor $(1-|m|/n)^{1/2}$ is a finite-sample correction and could be replaced by unity. Under certain conditions, $\hat{f}^{(0,1,0)}(\omega, 0, v)$ is consistent for $f^{(0,1,0)}(\omega, 0, v)$. 

14
Under $\mathbb{H}_0$, we have $\sigma^{(1,0)}_m(0, v) = 0$ for all $v \in \mathbb{R}^{d'}$ and all $m \neq 0$. Consequently, the generalized spectral derivative $f^{(0,1,0)}(\omega, 0, v)$ becomes a "flat spectrum" as a function of frequency $\omega$:

$$f^{(0,1,0)}_0(\omega, 0, v) \equiv \frac{1}{2\pi} \sigma^{(1,0)}_0(0, v) = \frac{1}{2\pi} \text{cov} \left( iZ_{t\Delta}, e^{ivZ_{t\Delta}} \right), \omega \in [-\pi, \pi] \text{ and } v \in \mathbb{R}^{d'}, \quad (3.14)$$

which can be consistently estimated by

$$\tilde{f}^{(0,1,0)}_0(\omega, 0, v) = \frac{1}{2\pi} \tilde{\sigma}^{(1,0)}_0(0, v), \omega \in [-\pi, \pi] \text{ and } v \in \mathbb{R}^{d'}.
\quad (3.15)$$

The estimators $\tilde{f}^{(0,1,0)}(\omega, 0, v)$ and $\tilde{f}^{(0,1,0)}_0(\omega, 0, v)$ converge to the same limit under $\mathbb{H}_0$ and generally converge to different limits under $\mathbb{H}_A$. Thus, any significant divergence between them is evidence of the violation of the MDS property and hence of the misspecification of the process. We can measure the distance between $\tilde{f}^{(0,1,0)}(\omega, 0, v)$ and $\tilde{f}^{(0,1,0)}_0(\omega, 0, v)$ by quadratic form:

$$\hat{Q} \equiv \frac{\pi n}{2} \int \int_{-\pi}^{\pi} \left\| \tilde{f}^{(0,1,0)}(\omega, 0, v) - \tilde{f}^{(0,1,0)}_0(\omega, 0, v) \right\|^2 d\omega dW(v) = \sum_{m=1}^{n-1} k^2(m/p)(n-m) \int \left\| \tilde{\sigma}^{(1,0)}_m(0, v) \right\|^2 dW(v),
\quad (3.16)$$

where the second equality follows by Parseval's identity and $W(v) = \prod_{c=1}^{d'} W_0(v_c)$ with $W_0 : \mathbb{R} \to \mathbb{R}^+$ a nondecreasing weighting function that weighs sets symmetric about the origin equally. Examples of $W_0(\cdot)$ include the CDF of any symmetric probability distribution, either discrete or continuous.

The proposed test statistic for the diffusion hypothesis is an appropriately standardized version of $\hat{Q}$,

$$\tilde{M}_0(p) = \left[ \sum_{m=1}^{n-1} k^2(m/p)(n-m) \int \left\| \tilde{\sigma}^{(1,0)}_m(0, v) \right\|^2 dW(v) - \tilde{C}_0(p) \right] / \sqrt{\tilde{D}_0(p)}$$

where

$$\tilde{C}_0(p) = \sum_{m=1}^{n-1} k^2(m/p) \frac{1}{n-m} \sum_{\tau=m+1}^{n-1} \left\| \tilde{Z}_{\tau \Delta} \right\|^2 \int \left| \tilde{\psi}_{(\tau-m)\Delta}(v) \right|^2 dW(v)$$

$$\tilde{D}_0(p) = 2 \sum_{m=1}^{n-2} k^2(m/p) \sum_{l=1}^{d'} k^2(l/p) \sum_{a=1}^{d'} \sum_{a'=1}^{d'} \int \int \right. \left. \left[ \tilde{Z}_{\tau \Delta} \tilde{Z}_{a' \Delta} \right] \tilde{\psi}_{(\tau-m)\Delta}(u) \tilde{\psi}^*_{(\tau-l)\Delta}(v) \right|^2 dW(u)dW(v),
\quad (3.17)$$

and $\tilde{\psi}_{\tau \Delta}(v) = e^{iv\tilde{Z}_{\tau \Delta}} - n^{-1} \sum_{\tau=1}^{n} e^{iv\tilde{Z}_{\tau \Delta}}$. Throughout, all unspecified integrals are taken on the support of $W(\cdot)$. The factors $\tilde{C}_0(p)$ and $\tilde{D}_0(p)$ are approximately the mean and the variance of the quadratic form $\hat{Q}$. The impact of conditional heteroskedasticity and other time-varying higher order conditional moments has already been taken into account. Note that $\tilde{M}_0(p)$ involves $d'$ and $2d'$-dimensional numerical integrations, which can be computationally cumbersome when $d'$ is large. In practice, one can use Monte Carlo simulation to approximate the integrals over $v$ and $u$. This can be obtained by using a large number of random draws
from the distribution $W(\cdot)$ and then computing the sample average as an approximation to the related integral. Such an approximation will be arbitrarily accurate provided the number of random draws is sufficiently large. Alternatively, we can use a nondecreasing step function $W(\cdot)$. This avoids numerical integration or Monte Carlo simulation, but the power of the test may be affected. In theory, the consistency property will not be preserved if only a finite number of grid points of $u$ and $v$ are used and the power of the test may depend on the choice of grid points of $u$ and $v$.

Before discussing the asymptotic properties of our test statistic, we summarize our testing procedure as follows: (1) Calculated the local polynomial estimator for the drift function and the realized volatility (covariance) for the quadratic variation (covariation); (2) Given the estimators in step (1), obtain the transformed processes $\{\widehat{Z}_{\tau\Delta}\}$ by the simple approximations to the Lebesgue integrals in (2.11); (3) Compute the test statistic $\widehat{M}_0(p)$ defined in (3.17); (4) Compare the value of $\widehat{M}_0(p)$ with the upper-tailed $N(0,1)$ critical value $C_\alpha$ at level $\alpha$ (this follows from Theorem 1 we shall prove in the next section). If $\widehat{M}_0(p) > C_\alpha$, then reject $H_0$ at level $\alpha$.

## 4 Asymptotic theory

### 4.1 Asymptotic distribution

To derive the null asymptotic distribution of the test statistic $\widehat{M}_0(p)$, the following regularity conditions are imposed.

**Assumption A.1:** Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. (i) The stochastic time series vector process $X_t = X_t(\omega)$, where $\omega \in \Omega$ and $t \in [0, T] \subset \mathbb{R}^+$, is a $d \times 1$ strictly stationary continuous-time Markov process with the marginal density $g(x)$, which is bounded and Lipschitz. Also, the joint density of $(X_1, X_t)$ is continuous and bounded by some constant independent of $l > 1$. (ii) A discrete sample $\{X_{\tau\Delta}\}_{\tau=1}^n$, where $\Delta$ is the sampling interval, is observed at equally spaced discrete times and $\{X_{\tau\Delta}\}_{\tau=1}^n$ is a $\beta$-mixing process with mixing coefficient $|\beta(l)| = CL^{-\nu}$ for some constant $\nu > 2$. (iii) $E \left| Z_{\tau\Delta} \right|^{8(1+\eta)} \leq C$ for some constant $\eta > 0$.

**Assumption A.2:** Let $m(x) \equiv E \left[ X_{(r+1)\Delta} - X_{r\Delta} | X_{r\Delta} = x \right]$. (i) The function $m(x)$ is $(r+1)-$th differentiable with respect to $x \in \mathbb{R}^d$ and $\frac{\partial^{(r+1)}}{\partial x^{(r+1)}} m(x)$ is Lipschitz of order $\alpha$: $\left| \frac{\partial^{(r+1)}}{\partial x^{(r+1)}} m(x_1) - \frac{\partial^{(r+1)}}{\partial x^{(r+1)}} m(x_2) \right| \leq l(|x_1 - x_2|^\alpha$, where $0 < \alpha \leq 1$ and $\int l^2(u) dW(u) < \infty$. (ii) $\sup_{x \in \mathbb{R}^d} \| m(x) \| \leq C$.

**Assumption A.3:** $[\widetilde{X}_t, \widetilde{X}_{t-\Delta}]_{t-\Delta} - [X_t, X_{t-\Delta}]_{t-\Delta} = O_P \left( \delta^{1/2} \right)$ and $[\widetilde{X}_t, \widetilde{X}_{t-\Delta}]_{t-\Delta} - [X_t, X_{t-\Delta}]_{t-\Delta} = O_P \left( \delta^{1/2} \right)$ for all $t$, where $[X_t, X_{t-\Delta}]_{t-\Delta} = \lim_{\delta \to 0} [X_t, X_{t-\Delta}]_{t-\Delta}$ and $[\widetilde{X}_t, \widetilde{X}_{t-\Delta}]_{t-\Delta} = \lim_{\delta \to 0} [\widetilde{X}_t, \widetilde{X}_{t-\Delta}]_{t-\Delta}$.

**Assumption A.4:** The function $K$ is a product kernel of some univariate kernel $K$, i.e., $K(u) = \prod_{j=1}^d K(u_j)$, where $K : \mathcal{G} \to \mathbb{R}^+$ is a symmetric and bounded function and $\mathcal{G}$ is a compact set. The function $H_j(u) \equiv u^j K(u)$ is Lipschitz for all $j$ with $0 \leq |j| \leq 2r + 1$.

**Assumption A.5:** (i) $k : \mathbb{R} \to [-1, 1]$ is a symmetric function that is continuous at zero and all points in $\mathbb{R}$ except for a finite number of points. (ii) $k(0) = 1$. (iii) $k(z) \leq c |z|^{-a}$ for some $a > \frac{3}{2}$ as $|z| \to \infty$. 

16
Assumption A.6: \( W : \mathbb{R}^d \to \mathbb{R}^+ \) is a nondecreasing weighting function that weighs sets symmetric about the origin equally, with \( \int \| u \|^4 dW(u) < \infty \).

Assumption A.1 imposes regularity conditions on the DGPs. Both univariate and multivariate continuous-time processes are covered. Following Ait-Sahalia (1996a, 1996b), Gallant and Long (1997), Gallant and Tauchen (1996), we impose regularity conditions on a discretely observed random sample. The \( \beta \)-mixing condition restricts the degree of temporal dependence in \( \{X_t\} \). We say that \( X_t \) is \( \beta \)-mixing if \( \beta(j) = \sup_{s \geq 1} E \left[ \sup_{A \in \mathcal{F}_{s+j}} |P(A\mid \mathcal{F}_s) - P(A)| \right] \to 0 \), as \( j \to \infty \), where \( \mathcal{F}_j \) is the \( \sigma \)-field generated by \( \{X_\tau : \tau = j, \ldots, s\} \), \( j \leq s \). Ait-Sahalia et al. (2009), Hjellvik et al. (1998) and Su and White (2007) also impose \( \beta \)-mixing conditions in related contexts. Suggested by Hansen and Scheinkman (1995) and Ait-Sahalia (1996a), one set of sufficient conditions for the \( \beta \)-mixing when \( d = 1 \) is: (i) \( \lim_{x \to l \text{ or } x \to u} \sigma(x, \theta) = 0 \); and (ii) \( \lim_{x \to -l \text{ or } x \to u} \frac{\sigma(x, \theta)}{\{2\mu(x, \theta) - \sigma(x, \theta) \partial \sigma(x, \theta)/\partial x\}} < \infty \), where \( l \) and \( u \) are left and right boundaries of \( X_t \) with possibly \( l = -\infty \) and/or \( u = +\infty \), and \( \pi(x, \theta) \) is the model-implied marginal density. Assumption A.2 provides conditions on the drift function. Assumption A.3 imposes conditions on the realized volatility and covariance. Under \( \mathbb{H}_0 \), \( [X^i, X^j]_{t-\Delta} \) and \( [X^i, X^j]_{t-\Delta} \) coincide with \( f^i_{t-\Delta} \sum_{k=1}^d \sigma_{i,k}(X_s) \sigma_{j,k}(X_s) ds \) and \( f^i_{t-\Delta} \sum_{k=1}^d \sigma_{i,k}(X_s) \sigma_{j,k}(X_s) ds \) respectively. Under \( \mathbb{H}_A \), they may take different forms, depending on the specific alternatives considered. See Section 5 for further discussions.

Assumption A.4 imposes regularity conditions on the kernel function used in local polynomial regression estimation. The same assumption has been imposed by Masry (1996a) and Ait-Sahalia et al. (2009). The condition on the boundedness and the compact support of \( K(\cdot) \) is imposed for the brevity of proofs and could be removed at the cost of a more tedious proof.\(^7\)

Assumption A.5 imposes regularity conditions on the kernel function \( k(\cdot) \) used for generalized cross-spectral estimation. This kernel is different from the kernel \( K(\cdot) \) used in the first stage nonparametric regression estimation of \( b(x) \). Here, \( k(\cdot) \) provides weighting for various lags, and is used to estimate the generalized spectral derivative \( f^{(0,1,0)}(\omega, 0, v) \). Among other things, the continuity of \( k(\cdot) \) at zero and \( k(0) = 1 \) ensures that the bias of the generalized spectral derivative \( \hat{f}^{(0,1,0)}(\omega, 0, v) \) vanishes to zero asymptotically as \( n \to \infty \). The condition on the tail behavior of \( k(\cdot) \) ensures that higher order lags will have little impact on the statistical properties of \( \hat{f}^{(0,1,0)}(\omega, 0, v) \). Assumption A.5 covers most commonly used kernels. For kernels with bounded support, such as the Bartlett and Parzen kernels, \( a = \infty \). For kernels with unbounded support, \( a \) is a finite positive real number. For example, \( a = 1 \) for the Daniell kernel \( k(z) = \sin(\pi z)/(\pi z) \), and \( a = 2 \) for the Quadratic-spectral kernel \( k(z) = 3/(\pi z)^2 \sin(\pi z)/(\pi z) - \cos(\pi z) \).

\(^7\)Alternatively, we could impose Hansen’s (2008) Assumption 3 on kernel functions, namely, for some \( \Lambda < \infty \) and \( L < \infty \), either \( K(u) = 0 \) for \( \|u\| > L \) and for all \( u, u' \in \mathbb{R}^d \), \( \|K(u) - K(u')\| \leq \Lambda \|u - u'\| \); or \( K(u) \) is differentiable, \( \|\partial /\partial u K(u)\| \leq \Lambda \|u\|^{-\nu} \) for \( \|u\| > L \), where \( \|u\| \equiv \max \{\|u_1\|, \ldots, \|u_d\|\} \). Here the kernel function is required to either have a truncated support and is Lipschitz or that it has a bounded derivative with an integrable tail. Our proof could go through with this assumption, but the trade-off is a strengthening requirement on the bandwidth. Since the choice of the bandwidth is more important than the choice of the kernel, and many commonly used kernels have compact support, we only consider the case of the compact support of the kernel \( K(u) \) in our formal analysis. Nevertheless, we examine the effect of allowing kernels with support on \( \mathbb{R}^d \) in our simulation study.
fourth moments satisfies Assumption A.6. Note that $W(\cdot)$ need not be continuous. This provides a convenient way to implement our tests, because we can avoid relatively high dimensional numerical integrations by using finitely many numbers of grid points for $u$ and $v$.

We now state the asymptotic distribution of the test statistic $\widehat{M}_0(p)$ under $\mathbb{H}_0$.

**Theorem 1:** Suppose that Assumptions A.1-A.6 hold, and $p = cn^\lambda$ for $c \in (0, \infty)$, $\lambda \in (\frac{1+n}{n}, (3 + \frac{1}{n-2})^{-1})$, $h \to 0$, $np^{1/2}\Delta^4 \to 0$, and $n\Delta h \to \infty$. Then under $\mathbb{H}_0$, $\widehat{M}_0(p) \to^d N(0, 1)$ as $n \to \infty$.

As an important feature of $\widehat{M}_0(p)$, the use of the estimated processes $\{\widehat{Z}_{\tau \Delta}\}$ in place of the true processes $\{Z_{\tau \Delta}\}$ has no impact on the limit distribution of $\widehat{M}_0(p)$. The reason is that the convergence rate of the drift function estimator and realized volatility is faster than that of the nonparametric kernel estimator to $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ to $f^{(0,1,0)}(\omega, 0, v)$. As a result, the limiting distribution of $\widehat{M}_0(p)$ is solely determined by $\widehat{Z}_{\tau \Delta}$ by $\widehat{Z}_{\tau \Delta}$ has no impact asymptotically. Of course, the low convergence rate may have impact on the finite sample performance of $\widehat{M}_0(p)$.

### 4.2 Asymptotic power

To gain insight into the nature of the alternatives that our test is able to detect, we examine the asymptotic behavior of $\widehat{M}_0(p)$ under $\mathbb{H}_A$. Define $b^*(x) = \lim_{\Delta \to 0} E\left[\frac{X_{(r+1)\Delta} - X_{\tau \Delta}}{\Delta} | X_{\tau \Delta} = x\right]$ and $Z_{\tau \Delta}^* \equiv Z_{\tau \Delta}(b^*(\cdot), [X^i, X^j], [X^i, X^j])$.

**Theorem 2:** Suppose Assumptions A.1-A.6 hold, and $p = cn^\lambda$ for $c \in (0, \infty)$, $\lambda \in (0, 1/2)$, $h \to 0$, $p\Delta^4 \to 0$, and $n\Delta h \to \infty$. Then under $\mathbb{H}_A$ and as $n \to \infty$,

\[
(p^{1/2}/n)\widehat{M}_0(p) \to^p \left[2D \int_0^\infty k^4(z)dz\right]^{-1/2} \int \left\|f^{(0,1,0)}(\omega, 0, v) - f^{(0,1,0)}_0(\omega, 0, v)\right\|^2 d\omega dW(v)
\]

\[
= \left[2D \int_0^\infty k^4(z)dz\right]^{-1/2} \sum_{m=1}^\infty \int \left\|\sigma_m^{(1,0)}(0, v)\right\|^2 dW(v)
\]

where

\[
D = 2 \sum_{a=1}^2 \sum_{a'=1}^2 E[Z^*_{ar} Z^*_{a'rr}] \int \int \int_{-\pi}^\pi |f(\omega, u, v)|^2 d\omega dW(u)dW(v), \tag{5.6}
\]

The constant $D$ takes into account the impact of the serial dependence in conditioning variables $\{e^{i \eta Z_{(\tau - m)\Delta}}\}$, which generally exists even under $\mathbb{H}_0$, due to the presence of the serial dependence in the conditional variance and higher order moments of $\{Z_{\tau \Delta}\}$. This differs from the i.i.d. case, where we can show that $D$ depends only on the marginal distribution of $\{Z_{\tau \Delta}\}$.

Suppose the autoregression function $E[Z_{\tau \Delta} \mid Z_{(\tau - m)\Delta}] \neq 0$ at some lag $m > 0$. Then we have $\int \left\|\sigma_m^{(1,0)}(0, v)\right\|^2 dW(v) > 0$ for any weighting function $W(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded support on $\mathbb{R}$. As a consequence, $\lim_{n \to \infty} P[\widehat{M}_0(p) > C(n)] = 1$ for any constant $C(n) = o(n/p^{1/2})$. Therefore, $\widehat{M}_0(p)$ has asymptotic unit power at any given significance level, whenever $E[Z_{\tau \Delta} \mid Z_{(\tau - m)\Delta}] \neq 0$ at some lag
The main advantage of $\widehat{M}_0(p)$ is that it can eventually detect all possible departure from the MDS property that causes $E[Z_{r\Delta} | Z_{(r-m)\Delta}] \neq 0$ at some lag $m \neq 0$. This avoids the blindness of searching for different alternatives when one has no prior information.

We also note that both $\hat{b}(\cdot)$ and $\overline{X, X}$ are used to compute $\widehat{M}_0(p)$ even under $\mathbb{H}_A$, i.e., when the DGP is not a diffusion process. Since our alternative model is a general non-diffusion Markov process, their respective limits, the instantaneous conditional mean $\lim_{\Delta \to 0} E \left[ \frac{X_{(r+1)\Delta} - X_{r\Delta}}{\Delta} | X_{r\Delta} = x \right]$ and the quadratic variation $[X, X]$, depend on the specific class of alternative models. For example, suppose the alternative model is a general jump-diffusion process

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + J_t,$$

where $J_t$ is a pure jump process, including Poisson and Compound Poisson processes for instance. Then by Bandi and Nguyen (2003) and Shreve (2004), we have

$$\lim_{\Delta \to 0} E \left[ \frac{X_{(r+1)\Delta} - X_{r\Delta}}{\Delta} | X_{r\Delta} = x \right] = b(x)$$

and

$$[X, X]^t_0 = \int_0^t \sigma^2(X_s)ds + \sum_{0 < s \leq t} (\Delta J_s)^2.$$

Therefore, the drift is still equal to the instantaneous conditional mean but the quadratic variation has an additional component due to the jumps.

5 Numerical Results

In this section, we shall examine the finite sample performance of our proposed testing procedure. As an empirical application, we also apply our test to both univariate time series including stock prices, interest rates, and foreign exchange rates, and bivariate time series such as LIBOR/Swap rates.

5.1 Monte Carlo Simulations

5.1.1 Simulation Design

We now study the finite sample performance of the testing procedure. The following three univariate diffusion models are considered to check the empirical size:

- DGP1: Vasicek (1977) Model
  $$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t$$
  with $(\kappa, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$.

- DGP2: CIR (Cox, Ross, and Ingersoll, 1985) Model
  $$dX_t = \kappa(\alpha - X_t)dt + \sigma \sqrt{X_t}dW_t$$
where \((\kappa, \alpha, \sigma^2) = (0.89218, 0.090495, 0.032742)\).

- DGP3: Inverse-Feller Model (Ahn and Gao, 1999):

\[
dX_t = X_t[\kappa - \beta X_t]dt + \sigma X_t^{3/2}dW_t
\]

where \((\kappa, \beta, \sigma^2) = (3.4387, 1.1361, 1.4209)\).

DGPs 1 and 2 have linear drift functions, where \(\alpha\) is the long run mean and \(\kappa\) is the speed of mean reversion. The smaller \(\kappa\) is, the stronger the serial dependence in \(\{X_t\}\) is, and consequently, the slower the process converges to the long run mean. The Vasicek model in DGP1 has constant volatility while the CIR model DGP2 has a state dependent diffusion function. The Inverse-Feller model in DGP3 proposed by Ahn and Gao(1999) is originally in the form of

\[
dX_t = X_t[\kappa - (\sigma^2 - \kappa \alpha) X_t]dt + \sigma X_t^{3/2}dW_t.
\]

We re-parameterize the model here for simplicity by setting \(\beta = \sigma^2 - \kappa \alpha\). From the simplified form employed here, we can see that the drift is essentially a quadratic function of \(X_t\) without a constant term. It is the quadratic term \(-\beta X_t^2\) that introduces the nonlinearity. The parameter values are the same as those in Pritsker (1998) and Hong and Li (2005).

We then consider the following univariate process to study the empirical power of the test \(\hat{M}_0(p)\).

- DGP4, Affine Jump-Diffusion model in Duffie, Pan and Singelton (2000):

\[
dX_t = \kappa(\alpha - X_t)dt + \sigma \sqrt{X_t} dW_t + JdN_t,
\]

where \(J\) is the random jump size which follows a \(N(\mu_J, \sigma_J^2)\) distribution, \(N_t\) is a Poisson process with arrival intensity \(\lambda\), \(W_t\) is a standard Brownian motion, the diffusion and jump processes are independent of each other and are also independent of jump size \(J\), \(\kappa = 0.89218\), \(\alpha = 0.090495\), \(\sigma^2 = 0.032742\), \(\sigma_J = 0.1\), and three different combinations of mean jump size and jump intensity:

- Case 1: \(\mu_J = 0.03\) and \(\lambda = 50\)
- Case 2: \(\mu_J = 0.05\) and \(\lambda = 50\)
- Case 3: \(\mu_J = 0.03\) and \(\lambda = 100\).

For the Jump-Diffusion model in DGP4, we assume that the coefficients are bounded and sufficient regularity conditions are satisfied so that a unique, strong solution exists (see Protter 2005 for details about the regularity conditions). DGP4 is actually a CIR model plus a jump term \(JdN_t\) which represents the "surprise news" in practice. Such a jump-diffusion model is still a Markov process but not a pure diffusion since the sample path is not continuous. Three cases are considered to check how our test \(\hat{M}_0(p)\) responds to different jump behaviors. For example, comparison between Case-1 and Case-2 reveals whether the power increases with the increase of the jump size and that between Case-1 and Case-3 can show the change of power if the jump intensity increases, i.e., more jumps are present in the process.
We now turn to bivariate continuous time processes, which are particularly interesting since our testing procedure may be the first test to check whether a multivariate process follows a diffusion. Ait-Sahalia’s (2002) and Kanaya’s (2007) are not applicable. Two bivariate diffusion models are considered to check the empirical size:

- **DGP5**: Bivariate Ornstein-Ulenbeck (O-U) model
  \[
  d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} d \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix},
  \]
  where \( W_{1t} \) and \( W_{2t} \) are two independent Brownian motions and \((\kappa_{11}, \kappa_{21}, \kappa_{22}, \sigma_{11}^2, \sigma_{22}^2) = (-0.1117, 1.1138, -1.1637, 0.000546, 0.002185)\).

- **DGP6**: Bivariate CIR model
  \[
  d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left( \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{bmatrix} \sigma_{11} \sqrt{X_{1t}} & 0 \\ 0 & \sigma_{22} \sqrt{X_{2t}} \end{bmatrix} d \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix},
  \]
  where \( W_{1t} \) and \( W_{2t} \) are two independent Brownian motions and \((\kappa_{11}, \kappa_{12}, \kappa_{21}, \kappa_{22}, \alpha_1, \alpha_2, \sigma_{11}^2, \sigma_{22}^2) = (-0.7, 0.3, 1.2, -0.8, 0.56, 0.48, 0.002, 0.001)\).

Both models are two-factor affine diffusion models used in the term structure literature (see Dai and Singleton 2000 and Ait-Sahalia and Kimmel 2009). For the Bivariate O-U model, \( \kappa_{21} \) controls the correlation between \( X_{1t} \) and \( X_{2t} \) and the transition density does not admit an explicit form unless \( \kappa_{21} = 0 \) under which it follows a Gaussian distribution (Duffee, 2002). For the Bivariate CIR model, the parameters specifying the correlation between \( X_{1t} \) and \( X_{2t} \) are \( \kappa_{21} \) and \( \kappa_{12} \), which have to be zero to allow for a closed-form transition density. In that case, it is equal to the product of two non-central chi-squared marginal distributions (Ait-Sahalia and Kimmel, 2009). The parameter values in the drift functions are the same as those in Song (2009).

Last, the following bivariate process is considered for checking the empirical power of the test:

- **DGP7**: Bivariate CIR Jump-Diffusion model
  \[
  d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left( \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{bmatrix} \sigma_{11} \sqrt{X_{1t}} & 0 \\ 0 & \sigma_{22} \sqrt{X_{2t}} \end{bmatrix} d \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix} + J dN_t,
  \]
  where \( W_{1t} \) and \( W_{2t} \) are two independent Brownian motions, \((\kappa_{11}, \kappa_{12}, \kappa_{21}, \kappa_{22}, \alpha_1, \alpha_2, \sigma_{11}^2, \sigma_{22}^2) = (-0.7, 0.3, 1.2, -0.8, 0.56, 0.48, 0.002, 0.001)\), \( J \) is the random jump size which follows a \( N(\mu_J, \Omega_J) \) distribution with \( \mu_J = (0.05, 0.01) \) and \( \Omega_J = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.12 \end{bmatrix} \), and \( N_t \) is a Poisson process with arrival intensity \( \lambda = 50 \).

DGP-7 is the Bivariate CIR model in DGP6 augmented by a jump term \( J dN_t \). The specification of the jump intensity implies that \( X_{1t} \) and \( X_{2t} \) jump together, i.e., the jump times are the same for the two components. However, the jump sizes are different and independent.
5.1.2 Simulation Results

For each DGP, we simulate 500 (the number of replications) data sets of a random sample \( \{X_t\}_{t=0}^{n} \) at the weekly frequency \((\Delta = 1/52)\) for \( n = 500, 1000 \) and 2000 respectively. These sample sizes correspond to around 10, 20 and 40 years of weekly data. Since our testing procedure requires the calculation of realized volatility and covariance for each weekly interval, we generate data sets at the daily frequency with \( M = 7 \), \( \delta = \Delta/M = 1/364 \) and 5 observations each day. We discard the first 4 data points and save the fifth as the daily observation for each day and then discard the first 6 data points and save the seventh as the weekly observation for each week.

An initial value is generated either by the known stationary distribution of \( X_t \) or by setting it equal to the average level of the interest rate data in Aït-Sahalia (1996b). We simulate data either from the transition density when it is available in closed-form like the Vasicek (1977) model or according to the so-called Euler scheme when the transition density is not analytic, like the Bivariate CIR model. To eliminate the effect of initial values, we generate 1000 more observations as the "burn-in" data points before starting the real simulation.

To calculate the test statistic \( \hat{M}_0(p) \), we first obtain the local linear estimator \( \hat{b}(\cdot) \) of the drift function in (3.6) with \( r = 1 \) and the realized volatility and covariance \( [X^t, X^t]^\top_{t-\Delta} \) and \( [X^t, X^t]^\top_{t-\Delta} \). Similar to Kanaya (2007), the bandwidth \( h \) for \( \hat{b}(\cdot) \) is chosen as \( h = 4\hat{\sigma}_X n^{-1/4.5} \) where \( \hat{\sigma}_X \) is the standard deviation of the observations. The Bartlett kernel is used in computing the test statistic \( \hat{M}_0(p) \). And we choose the standard multivariate normal CDF \( \Phi(\cdot) \) as the weighting distribution \( W(\cdot) \). To reduce the computational costs in the simulation study, we generate \( u \) and \( v \) from \( \Phi(\cdot) \) with each \( u \) and \( v \) having 30 symmetric grid points respectively. Our simulation experience indicates that choices of kernel function \( k(\cdot) \) and weighting function \( W(\cdot) \) have no substantial impact on both the size and power of the tests. Like Hong (1999), we use a data-driven \( \hat{p}_0 \) via a plug-in method that minimizes the asymptotic integrated MSE of the generalized spectral derivative estimator \( \hat{J}^{(0,1,0)} \) with the Bartlett kernel. To examine the sensitivity of the choice of a preliminary bandwidth \( \hat{p} \) on the size and power of the \( \hat{M}_0(p) \) test, we consider \( \hat{p} \) in the range of 10 to 40. We use the Gaussian kernel for \( K(\cdot) \) in the local linear estimation.

Table I and III report the empirical sizes of the \( \hat{M}_0(p) \) test at the 10% and 5% levels under univariate DGPs 1-3 and bivariate DGPs 5-6 respectively. For the three univariate DGPs, \( \hat{M}_0(p) \) tends to underreject a bit when \( n = 500 \) but is the underrejection is not excessive. It improves very quickly as the sample sizes increase. When \( n = 2000 \), the empirical levels are very close to the nominal levels. For the two bivariate DGPs, \( \hat{M}_0(p) \) tends to overreject when \( n = 500 \) but the overrejection is not excessive either. Similar to univariate cases, the empirical levels improve with \( n \) and are close to the nominal levels when \( n = 2000 \).

Table II and III report the empirical powers of the \( \hat{M}_0(p) \) test for DGPs 4 and 7 respectively. Due to the jump terms which make the sample path discontinuous, neither of these two DGPs is a diffusion although

---

8 We first generate 15 grid points \( u_0 \) and \( v_0 \) from \( \Phi(\cdot) \) and obtain \( u' = [u_0' - u_0] \) and \( v' = [v_0' - v_0] \) to ensure symmetry. Preliminary experiments with different numbers of grid points show that simulation results are not very sensitive to the choice of numbers. Concerned with the computational cost in the simulation study, we are satisfied with current results with 30 grid points.

9 We have tried Parzen kernel and obtain similar results (not reported here).
both are still Markov. For all three cases of DGP4, the $\hat{M}_0(p)$ test has reasonable power, with rejection rates increased from around 20% to around 40% at the 5% level when $n$ increases from 500 to 2000. As expected, the empirical power increases with the jump size and jump intensity. For the bivariate DGP7, the empirical power of our $\hat{M}_0(p)$ test may seem a bit low. But the empirical rejection rates indeed increase with $n$.

In summary, we observe: (1) Unlike other nonparametric tests in the literature, the $\hat{M}_0(p)$ test has good sizes in finite samples for both univariate and bivariate processes. No bootstrap procedure is needed. (2) The test has reasonable powers against the non-diﬀusion alternatives. (3) Sizes and powers of $\hat{M}_0(p)$ are not very sensitive to the choice of the preliminary lag order $\hat{p}$ for all DGPs considered.

5.2 Applications to Financial Data

As documented by Ait-Sahalia (1996a) and Hong and Li (2005), many popular diﬀusion models for spot interest rates such as Vasicek (1977), Cox, Ingersoll and Ross (1985), Chan, Karolyi, Longstaﬀ and Sanders (1992), Ait-Sahalia (1996a) and Ahn and Gao (1999) are all strongly rejected with real interest rate data. Since these models are all parametric diﬀusion models, the rejection could be due to the misspecification of the parametric forms or the violation of the diﬀusion assumption. If the diﬀusion assumption is valid, we can still remain in the diﬀusion framework and search for more ﬂexible speciﬁcations of the drift and diﬀusion functions. Otherwise, we should turn to models with other salient characteristics such as jumps. In fact, the extensive literature on jumps such as Andersen et al. (2002), Pan (2002), Ait-Sahalia (2004), Johannes (2004), and Piazessi (2005) and Ait-Sahalia and Jacod (2008) suggests that the diﬀusion hypothesis is not appropriate for many ﬁnancial variables. In this section, we shall apply our test to formally check whether some ﬁnancial time series follow diﬀusion processes. Our test addresses the diﬀusion hypothesis directly and answers the question that whether the diﬀusion processes are adequate for modelling those ﬁnancial variables. The tests for jumps like Ait-Sahalia and Jacod (2008) can also shed some light on this question by pointing out whether the sample path is discontinuous or not. However, it is possible in practice that jumps, although present, have negligible impact on the DGPs and the diﬀusion models are still able to capture the dynamics of the time series. Therefore, our test is more reliable in determining whether diﬀusion processes should be employed in ﬁnancial modelling.

We apply our test to three important univariate ﬁnancial time series, i.e., S&P 500 stock price index, 7-day Eurodollar rate, and Japanese Yen exchange rate, and bivariate yields series on ordinary, ﬁxed-for-variable rate U.S. dollar swap contracts (see Dai and Singleton 2000 and Piazessi 2010 for detailed discussions about the swap rates), obtained from Datastream. The joint dynamics of swap yields with diﬀerent maturities have been the focus of term structure models of interest rates for over three decades due to the importance of bond yields for forecasting, monetary policy, debt policy, and derivative pricing and hedging (see Section 1.2 of Piazzesi 2009). Affine term structure models (ATSMs), including many previously proposed models by Vasicek (1977), Cox, Ingersoll and Ross (1985), Longstaﬀ and Schwartz (1992), Chen and Scott (1992), and so on (see Duffie and Kan (1996) for more examples) as special cases, are very popular among both practitioners and academics because of its analytic tractability. Duffie and Kan (1996) lay down the general framework of ATSMs by summarizing the primitive assumptions while Dai and Singleton (2000) analyze the admissibility
conditions to characterize the maximally flexible and empirically identifiable ATSMs driven by pure diffusions. Nowadays, it becomes standard to employ Dai and Singleton (2000) framework for empirical studies in term structure literature; see, for example, Duffee (2002), Cheridito et al. (2007), Thompson(2008), Egorov, Li, and Ng (2008), and Brandt and He (2006).

The three univariate time series are all daily data from January 1, 1988 to December 31, 2006 and the swap rates are also daily but from August 13, 1990 to December 31, 2008, containing maturities of 2-year, 3-year, 4-year, 5-year, 7-year and 10-year. We also generate weekly series for all the data considered by selecting Wednesday series (if a Wednesday is a holiday then the previous Tuesday is used), with the three univariate series have 991 and the swap rates 950 observations respectively. The use of weekly data has the advantage of avoiding the so-called weekend effect and other biases associated with nontrading and asynchronous rates, which often appear in higher frequency data. For the three univariate series, we consider two subsamples, January 1, 1988 to December 31, 1997 and January 1, 1998 to December 31, 2006, to examine the sensitivity of our conclusion to the possible structural changes (see Chen and Hong 2010 for some tests for structural changes applied to these univariate series). Furthermore, as is a standard practice, we use S&P 500 log returns, 7-day Eurodollar rate changes, Japanese Yen log returns, and swap rate changes to avoid the possible impact of unit root in the level series.

Tables IV and V report the test statistics and p-values of our $\hat{M}_0(p)$ test for the three univariate financial time series at both the daily and weekly frequency. For daily data, we set $\Delta = 1/52$ and $M = 5$ and for weekly data, $\Delta = 1/12$ and $M = 4$. For all the data and sample periods considered, our test is quite robust to the preliminary bandwidth choice of $\Delta$, ranging from 16 to 25. At both weekly and monthly frequencies, we find strong evidences against the diffusion hypothesis for S&P 500 returns, 7-day Eurodollar rate changes and Japanese Yen returns, for both the whole sample and the two subsamples: the p-values of our test are smaller than 5% except that for the weekly 7-day Eurodollar rate changes, the p-values are 5.95% at most.

Table VI reports the test statistics and p-values of our $\hat{M}_0(p)$ test for two bivariate swap rate changes series: 2-year with 3-year maturities swap rates and 2-year with 10-year maturities swap rates\(^{10}\) at both daily and weekly frequency. Similar to the univariate series, strong evidences are found against the diffusion hypothesis at both frequencies: the p-values of our test are smaller than 3%. This shows directly that multivariate affine diffusion term structure models, although very popular in modelling the term structure dynamics (Dai and Singleton 2000, Piazessi, 2010), are not adequate for capturing the dynamics of swap rates. Our results confirms some concerns in the term structure literature that "state variables with jumps have received relatively less attention in the empirical literature on DTSMs" (Dai and Singleton 2003) and "there is little work of jump-diffusion term structure models" (Johannes and Polson 2009).

\(^{10}\)We also try other combinations of 2-year, 3-year, 4-year, 5-year, 7-year and 10-year swap rates. The results are very similar and available from the authors upon request.
6 Conclusion

We develop a nonparametric test to check whether the underlying continuous time process is a diffusion, i.e., whether a process can be represented by a SDE. Our testing procedure utilizes the infinitesimal operator based martingale characterization of diffusion models, under which the null hypothesis is equivalent to a martingale difference property of the transformed processes. Then a generalized spectral derivative test is applied to check the martingale property, where the drift function is estimated via local polynomial regression and the diffusion function is integrated out by quadratic variation. Such a testing procedure is feasible and convenient because the infinitesimal operator of the diffusion process, unlike the transition density, has a closed-form expression of the drift and diffusion functions. The proposed test is applicable to both univariate and multivariate continuous time processes and has a $N(0, 1)$ limit distribution. Simulation studies show that the proposed test has reasonable size and all-around power against non-diffusion alternatives in finite samples. We apply our test to several important financial time series and find some evidence that the diffusion hypothesis may not be suitable for many financial time series.
References


<table>
<thead>
<tr>
<th></th>
<th>$n = 500$</th>
<th>$n = 1000$</th>
<th>$n = 2000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>0.062</td>
<td>0.116</td>
<td>0.104</td>
</tr>
<tr>
<td>5%</td>
<td>0.042</td>
<td>0.086</td>
<td>0.060</td>
</tr>
</tbody>
</table>

**DGP1-Vasicek Model**

<table>
<thead>
<tr>
<th></th>
<th>$n = 500$</th>
<th>$n = 1000$</th>
<th>$n = 2000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>0.064</td>
<td>0.110</td>
<td>0.104</td>
</tr>
<tr>
<td>5%</td>
<td>0.040</td>
<td>0.084</td>
<td>0.058</td>
</tr>
</tbody>
</table>

**DGP2-CIR Model**

<table>
<thead>
<tr>
<th></th>
<th>$n = 500$</th>
<th>$n = 1000$</th>
<th>$n = 2000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>0.084</td>
<td>0.120</td>
<td>0.098</td>
</tr>
<tr>
<td>5%</td>
<td>0.064</td>
<td>0.078</td>
<td>0.058</td>
</tr>
</tbody>
</table>

**DGP3-Inverse Feller Model**

<table>
<thead>
<tr>
<th></th>
<th>$n = 500$</th>
<th>$n = 1000$</th>
<th>$n = 2000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>0.070</td>
<td>0.092</td>
<td>0.106</td>
</tr>
<tr>
<td>5%</td>
<td>0.036</td>
<td>0.048</td>
<td>0.052</td>
</tr>
</tbody>
</table>

Notes:

(i): 500 iterations;

(ii): DGPs1-3 are Vasicek model, $dX_t = \kappa(\alpha-X_t)dt + \sigma dW_t$ with $(\kappa, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$. CIR model $dX_t = \kappa(\alpha-X_t)dt + \kappa\sqrt{X_t}dW_t$ where $(\kappa, \alpha, \sigma^2) = (0.89218, 0.090495, 0.032742)$ and Inverse-Feller model $dX_t = (\kappa-X_t -\beta X_t^2)dt + \sigma X_t^{3/2}dW_t$ where $(\kappa, \beta, \sigma^2) = (3.4387, 1.1361, 1.4209)$

(iii): The bandwidth $h$ for the local linear estimator of the drift function is chosen as $4\hat{S}_X n^{-1/4.5}$, where $\hat{S}_X$ is the sample standard deviation of the observations. The Gaussian kernel is used as $K(\cdot)$ for this local linear estimator.

(iv): $\bar{p}$, the preliminary bandwidth, is used in a plug-in method to choose a data-dependent bandwidth $\hat{p}_0$ with the Bartlett kernel is used. The Bartlett kernel is also used for computing $\hat{M}_0(\hat{p}_0)$. 
### TABLE II

**Empirical Power Under Univariate DGP- 4**

<table>
<thead>
<tr>
<th></th>
<th>n = 500</th>
<th></th>
<th>n = 1000</th>
<th></th>
<th>n = 2000</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td><strong>( \bar{p} )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.192</td>
<td>0.144</td>
<td>0.256</td>
<td>0.202</td>
<td>0.394</td>
<td>0.348</td>
</tr>
<tr>
<td>20</td>
<td>0.194</td>
<td>0.144</td>
<td>0.256</td>
<td>0.202</td>
<td>0.396</td>
<td>0.348</td>
</tr>
<tr>
<td>30</td>
<td>0.190</td>
<td>0.146</td>
<td>0.264</td>
<td>0.206</td>
<td>0.396</td>
<td>0.350</td>
</tr>
<tr>
<td>40</td>
<td>0.190</td>
<td>0.148</td>
<td>0.260</td>
<td>0.206</td>
<td>0.398</td>
<td>0.350</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DGP4-Affine Jump Diffusion Model with ( \mu_J = 0.3 ) and ( \lambda = 50 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.268</td>
<td>0.210</td>
<td>0.384</td>
<td>0.282</td>
<td>0.518</td>
<td>0.452</td>
</tr>
<tr>
<td>20</td>
<td>0.268</td>
<td>0.210</td>
<td>0.384</td>
<td>0.290</td>
<td>0.512</td>
<td>0.450</td>
</tr>
<tr>
<td>30</td>
<td>0.262</td>
<td>0.212</td>
<td>0.382</td>
<td>0.294</td>
<td>0.510</td>
<td>0.446</td>
</tr>
<tr>
<td>40</td>
<td>0.262</td>
<td>0.214</td>
<td>0.382</td>
<td>0.302</td>
<td>0.516</td>
<td>0.444</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DGP4-Affine Jump Diffusion Model with ( \mu_J = 0.6 ) and ( \lambda = 50 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.190</td>
<td>0.146</td>
<td>0.312</td>
<td>0.262</td>
<td>0.460</td>
<td>0.424</td>
</tr>
<tr>
<td>20</td>
<td>0.198</td>
<td>0.150</td>
<td>0.316</td>
<td>0.260</td>
<td>0.462</td>
<td>0.426</td>
</tr>
<tr>
<td>30</td>
<td>0.198</td>
<td>0.152</td>
<td>0.310</td>
<td>0.268</td>
<td>0.460</td>
<td>0.420</td>
</tr>
<tr>
<td>40</td>
<td>0.196</td>
<td>0.154</td>
<td>0.308</td>
<td>0.260</td>
<td>0.468</td>
<td>0.424</td>
</tr>
</tbody>
</table>

Notes:  
(i): 500 iterations;  
(ii): DGP-4 is an affine jump-diffusion model, \( dX_t = \kappa(\alpha - X_t)dt + \sigma \sqrt{X_t}dW_t + JdN_t \), where \( J \) is the random jump size which follows a \( N(\mu_J, \sigma_J^2) \) distribution, and \( N_t \) is a Poisson process with arrival intensity \( \lambda \). The parameter values are \( \kappa = 0.89218, \alpha = 0.090495, \sigma = 0.032742, \mu_J = 0.3 \) or 0.5, \( \sigma_J = 0.1 \) and \( \lambda = 50 \) or 100.  
(iii): The bandwidth \( h \) for the local linear estimator of the drift function is chosen as \( 4\hat{S}_X n^{-1/4.5} \), where \( \hat{S}_X \) is the sample standard deviation of the observations. The Gaussian kernel is used as \( K(\cdot) \) for this local linear estimator.  
(iv): \( \bar{p} \), the preliminary bandwidth, is used in a plug-in method to choose a data-dependent bandwidth \( \hat{p}_0 \) with the Bartlett kernel is used. The Bartlett kernel is also used for computing \( \widehat{M}_0(\hat{p}_0) \).
TABLE III

<table>
<thead>
<tr>
<th>Empirical Size and Power Under Multivariate DGPs 5-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 500 )</td>
</tr>
<tr>
<td>( \bar{p} )</td>
</tr>
<tr>
<td>10%</td>
</tr>
<tr>
<td>Empirical Size under DGP5-Bivariate Feller Model</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>Empirical Size under DGP6-Bivariate CIR Model</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>Empirical Power under DGP7-Bivariate CIR Jump Diffusion Model</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>40</td>
</tr>
</tbody>
</table>

Notes: (i): 500 iterations; (ii): DGP5 is the Bivariate O-U

\[
    d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}
\]

with \((\kappa_{11}, \kappa_{21}, \kappa_{22}, \sigma_{11}^2, \sigma_{22}^2) = (-0.1117, 1.1138, -1.1637, 0.000546, 0.002185)\); DGP6 is the Bivariate CIR

\[
    d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix}\right) dt + \begin{bmatrix} \sigma_{11}\sqrt{X_{1t}} & 0 \\ 0 & \sigma_{22}\sqrt{X_{2t}} \end{bmatrix} \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}
\]

with \((\kappa_{11}, \kappa_{12}, \kappa_{21}, \kappa_{22}, \alpha_1, \alpha_2, \sigma_{11}^2, \sigma_{22}^2) = (-0.7, 0.3, 1.2, -0.8, 0.56, 0.48, 0.002, 0.001)\);

(iii): DGP7 is Bivariate CIR Jump Diffusion model which is the Bivariate CIR model in DGP6 augmented by a jump term \(JdN_t\) where \(J\) is the random jump size which follows a \(N(\mu_j, \Omega_j)\) distribution with \(\mu_j = (0.05, 0.01)\) and \(\Omega_j = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.12 \end{bmatrix}\), and \(N_t\) is a Poisson process with arrival intensity \(\lambda = 50\); and \(W_{1t}\) and \(W_{2t}\) are two independent Brownian motions

(iv): The bandwidth \(h\) for the local linear estimator of the drift function is chosen as \(4\hat{S}_X n^{-1/6}\), where \(\hat{S}_X\) is the sample standard deviation of the observations. The Gaussian kernel is used as \(K(\cdot)\) for this local linear estimator.

(v): \(\bar{p}\), the preliminary bandwidth, is used in a plug-in method to choose a data-dependent bandwidth \(b_0\) with the Bartlett kernel is used. The Bartlett kernel is also used for computing \(\hat{M}_0(\hat{p}_0)\).
TABLE IV: Diffusion test for S&P500, interest rate and exchange rate at Daily Frequency

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P 500</th>
<th>7-day Eurodollar rate</th>
<th>Japanese Yen</th>
<th></th>
<th>7-day Eurodollar rate</th>
<th>Japanese Yen</th>
<th></th>
<th>7-day Eurodollar rate</th>
<th>Japanese Yen</th>
<th></th>
<th>7-day Eurodollar rate</th>
<th>Japanese Yen</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>47.1889</td>
<td>0.0000</td>
<td>9.1372</td>
<td>0.0000</td>
<td>41.3395</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>41.8184</td>
<td>0.0000</td>
<td>9.1002</td>
<td>0.0000</td>
<td>40.6253</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>41.5180</td>
<td>0.0000</td>
<td>9.0641</td>
<td>0.0000</td>
<td>40.0251</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>41.2322</td>
<td>0.0000</td>
<td>9.0155</td>
<td>0.0000</td>
<td>39.5066</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>40.9648</td>
<td>0.0000</td>
<td>8.9432</td>
<td>0.0000</td>
<td>39.0449</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>40.7163</td>
<td>0.0000</td>
<td>8.8576</td>
<td>0.0000</td>
<td>38.6019</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>40.4802</td>
<td>0.0000</td>
<td>8.7696</td>
<td>0.0000</td>
<td>38.1683</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>40.2486</td>
<td>0.0000</td>
<td>8.6842</td>
<td>0.0000</td>
<td>37.7524</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>40.0207</td>
<td>0.0000</td>
<td>8.6035</td>
<td>0.0000</td>
<td>37.3572</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>39.7978</td>
<td>0.0000</td>
<td>8.5282</td>
<td>0.0000</td>
<td>36.9851</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>01/01/1988-12/31/1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>14.0152</td>
<td>0.0000</td>
<td>5.2220</td>
<td>0.0000</td>
<td>28.8712</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>14.0029</td>
<td>0.0000</td>
<td>5.1959</td>
<td>0.0000</td>
<td>28.2610</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>13.9885</td>
<td>0.0000</td>
<td>5.1706</td>
<td>0.0000</td>
<td>27.6747</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>13.9725</td>
<td>0.0000</td>
<td>5.1483</td>
<td>0.0000</td>
<td>27.0218</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>13.9531</td>
<td>0.0000</td>
<td>5.1166</td>
<td>0.0000</td>
<td>26.3225</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>13.9301</td>
<td>0.0000</td>
<td>5.0659</td>
<td>0.0000</td>
<td>25.6600</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>13.9032</td>
<td>0.0000</td>
<td>5.0074</td>
<td>0.0000</td>
<td>25.0592</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>13.8714</td>
<td>0.0000</td>
<td>4.9486</td>
<td>0.0000</td>
<td>24.5220</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>13.8329</td>
<td>0.0000</td>
<td>4.8928</td>
<td>0.0000</td>
<td>24.0395</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>13.7868</td>
<td>0.0000</td>
<td>4.8410</td>
<td>0.0000</td>
<td>23.6061</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>01/01/1998-12/31/2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>27.0102</td>
<td>0.0000</td>
<td>2.4555</td>
<td>0.0000</td>
<td>31.8021</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>26.7726</td>
<td>0.0000</td>
<td>2.4542</td>
<td>0.0000</td>
<td>31.4849</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>26.5438</td>
<td>0.0000</td>
<td>2.4529</td>
<td>0.0000</td>
<td>31.1422</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>26.3235</td>
<td>0.0000</td>
<td>2.4515</td>
<td>0.0000</td>
<td>30.7762</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>26.1087</td>
<td>0.0000</td>
<td>2.4501</td>
<td>0.0000</td>
<td>30.4208</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>25.9027</td>
<td>0.0000</td>
<td>2.4487</td>
<td>0.0000</td>
<td>30.0818</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>25.7081</td>
<td>0.0000</td>
<td>2.4472</td>
<td>0.0000</td>
<td>29.7383</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>25.5065</td>
<td>0.0000</td>
<td>2.4456</td>
<td>0.0000</td>
<td>29.3910</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>25.2919</td>
<td>0.0000</td>
<td>2.4438</td>
<td>0.0000</td>
<td>29.0544</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>25.0640</td>
<td>0.0000</td>
<td>2.4420</td>
<td>0.0000</td>
<td>28.7367</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: (i): The bandwidth $h$ for the local linear estimator of the drift function is chosen as $4\hat{S}_X n^{-1/4.5}$, where $\hat{S}_X$ is the sample standard deviation of the observations. The Gaussian kernel is used as $K(\cdot)$ for this local linear estimator. (ii): $\bar{p}$, the preliminary bandwidth, is used in a plug-in method to choose a data-dependent bandwidth $\hat{p}_0$ with the Bartlett kernel is used. The Bartlett kernel is also used for computing $\hat{M}_0(\hat{p}_0)$.  

34
### TABLE V: Diffusion test for S&P500, interest rate and exchange rate at Weekly Frequency

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P 500</th>
<th>7-day Eurodollar rate</th>
<th>Japanese Yen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-values</td>
<td>Statistics</td>
<td>p-values</td>
</tr>
<tr>
<td>01/01/1988-12/31/2006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.0000</td>
<td>7.14562</td>
<td>0.0074</td>
</tr>
<tr>
<td>17</td>
<td>0.0000</td>
<td>7.03430</td>
<td>0.0047</td>
</tr>
<tr>
<td>18</td>
<td>0.0000</td>
<td>6.91957</td>
<td>0.0031</td>
</tr>
<tr>
<td>19</td>
<td>0.0000</td>
<td>6.80966</td>
<td>0.0024</td>
</tr>
<tr>
<td>20</td>
<td>0.0000</td>
<td>6.70099</td>
<td>0.0022</td>
</tr>
<tr>
<td>21</td>
<td>0.0000</td>
<td>6.59507</td>
<td>0.0021</td>
</tr>
<tr>
<td>22</td>
<td>0.0000</td>
<td>6.50142</td>
<td>0.0018</td>
</tr>
<tr>
<td>23</td>
<td>0.0000</td>
<td>6.42191</td>
<td>0.0016</td>
</tr>
<tr>
<td>24</td>
<td>0.0000</td>
<td>6.36015</td>
<td>0.0015</td>
</tr>
<tr>
<td>25</td>
<td>0.0000</td>
<td>6.31677</td>
<td>0.0014</td>
</tr>
<tr>
<td>01/01/1988-12/31/1997</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.0000</td>
<td>12.5432</td>
<td>0.0595</td>
</tr>
<tr>
<td>17</td>
<td>0.0000</td>
<td>12.3155</td>
<td>0.0590</td>
</tr>
<tr>
<td>18</td>
<td>0.0000</td>
<td>12.0886</td>
<td>0.0400</td>
</tr>
<tr>
<td>19</td>
<td>0.0000</td>
<td>11.8711</td>
<td>0.0220</td>
</tr>
<tr>
<td>20</td>
<td>0.0000</td>
<td>11.6676</td>
<td>0.0132</td>
</tr>
<tr>
<td>21</td>
<td>0.0000</td>
<td>11.4773</td>
<td>0.0075</td>
</tr>
<tr>
<td>22</td>
<td>0.0000</td>
<td>11.2985</td>
<td>0.0050</td>
</tr>
<tr>
<td>23</td>
<td>0.0000</td>
<td>11.1310</td>
<td>0.0033</td>
</tr>
<tr>
<td>24</td>
<td>0.0000</td>
<td>10.9758</td>
<td>0.0024</td>
</tr>
<tr>
<td>25</td>
<td>0.0000</td>
<td>10.8322</td>
<td>0.0022</td>
</tr>
<tr>
<td>01/01/1998-12/31/2006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.0000</td>
<td>5.2488</td>
<td>0.0333</td>
</tr>
<tr>
<td>17</td>
<td>0.0000</td>
<td>5.1356</td>
<td>0.0333</td>
</tr>
<tr>
<td>18</td>
<td>0.0000</td>
<td>5.0272</td>
<td>0.0333</td>
</tr>
<tr>
<td>19</td>
<td>0.0000</td>
<td>4.9236</td>
<td>0.0333</td>
</tr>
<tr>
<td>20</td>
<td>0.0000</td>
<td>4.8187</td>
<td>0.0323</td>
</tr>
<tr>
<td>21</td>
<td>0.0000</td>
<td>4.7099</td>
<td>0.0319</td>
</tr>
<tr>
<td>22</td>
<td>0.0000</td>
<td>4.5968</td>
<td>0.0332</td>
</tr>
<tr>
<td>23</td>
<td>0.0000</td>
<td>4.4688</td>
<td>0.0345</td>
</tr>
<tr>
<td>24</td>
<td>0.0000</td>
<td>4.3192</td>
<td>0.0368</td>
</tr>
<tr>
<td>25</td>
<td>0.0000</td>
<td>4.1509</td>
<td>0.0423</td>
</tr>
</tbody>
</table>

Notes: (i): The bandwidth \( h \) for the local linear estimator of the drift function is chosen as \( 4\hat{S}_X n^{-1/4.5} \), where \( \hat{S}_X \) is the sample standard deviation of the observations. The Gaussian kernel is used as \( K(\cdot) \) for this local linear estimator. (ii): \( \hat{p} \), the preliminary bandwidth, is used in a plug-in method to choose a data-dependent bandwidth \( \hat{p}_0 \) with the Bartlett kernel is used. The Bartlett kernel is also used for computing \( \hat{M}_0(\hat{p}_0) \).
<table>
<thead>
<tr>
<th></th>
<th>Daily Data</th>
<th></th>
<th>Weekly Data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2y&amp;3y Statistics</td>
<td>p-values</td>
<td>2y&amp;10y Statistics</td>
<td>p-values</td>
</tr>
<tr>
<td>16</td>
<td>23.9632</td>
<td>0.0000</td>
<td>31.0775</td>
<td>0.0000</td>
</tr>
<tr>
<td>17</td>
<td>23.3109</td>
<td>0.0000</td>
<td>29.9696</td>
<td>0.0000</td>
</tr>
<tr>
<td>18</td>
<td>22.6740</td>
<td>0.0000</td>
<td>28.9501</td>
<td>0.0000</td>
</tr>
<tr>
<td>19</td>
<td>22.0501</td>
<td>0.0000</td>
<td>28.0160</td>
<td>0.0000</td>
</tr>
<tr>
<td>20</td>
<td>21.4609</td>
<td>0.0000</td>
<td>27.1607</td>
<td>0.0000</td>
</tr>
<tr>
<td>21</td>
<td>20.8999</td>
<td>0.0000</td>
<td>26.3849</td>
<td>0.0000</td>
</tr>
<tr>
<td>22</td>
<td>20.3716</td>
<td>0.0000</td>
<td>25.6741</td>
<td>0.0000</td>
</tr>
<tr>
<td>23</td>
<td>19.8803</td>
<td>0.0000</td>
<td>24.9942</td>
<td>0.0000</td>
</tr>
<tr>
<td>24</td>
<td>19.4268</td>
<td>0.0000</td>
<td>24.3368</td>
<td>0.0000</td>
</tr>
<tr>
<td>25</td>
<td>18.9864</td>
<td>0.0000</td>
<td>23.7003</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Notes: (i): The bandwidth $h$ for the local linear estimator of the drift function is chosen as $4\hat{S}_X n^{-1/6}$, where $\hat{S}_X$ is the sample standard deviation of the observations. The Gaussian kernel is used as $K(\cdot)$ for this local linear estimator.

(ii): $\hat{\eta}$, the preliminary bandwidth, is used in a plug-in method to choose a data-dependent bandwidth $\hat{p}_0$ with the Bartlett kernel is used. The Bartlett kernel is also used for computing $\hat{M}_0(\hat{p}_0)$.