A Martingale Approach for Testing Diffusion Models Based on Infinitesimal Operator

Zhaogang Song*
Department of Economics, Cornell University

First Draft: November 2008
This Draft: February 2009

Abstract: I develop an omnibus specification test for diffusion models based on the infinitesimal operator instead of the already extensively used transition density. The infinitesimal operator-based identification of the diffusion process is equivalent to a "martingale hypothesis" for the new processes transformed from the original diffusion process. The transformation is via the celebrated "martingale problems". My test procedure is to check the "martingale hypothesis" via a multivariate generalized spectral derivative based approach which enjoys many good properties. The infinitesimal operator of the diffusion process enjoys the nice property of being a closed-form expression of drift and diffusion terms. This makes my test procedure capable of checking both univariate and multivariate diffusion models and particularly powerful and convenient for the multivariate case. In contrast checking the multivariate diffusion models is very difficult by transition density-based methods because transition density does not have a closed-form in general. Moreover, different transformed martingale processes contain different separate information about the drift and diffusion terms and their interactions. This motivates us to suggest a separate inference-based test procedure to explore the sources when rejection of a parametric form happens. Finally, simulation studies are presented and possible future researches using the infinitesimal operator-based martingale characterization are discussed.

Keywords: Diffusion; Jump Diffusion; Markov; Martingale; Martingale problem; Semi-group; Drift; Infinitesimal operator; Transition density; Generalized spectrum.

*Correspondence: Zhaogang Song, Uris Hall 445, Department of Economics, Cornell University, Ithaca, NY 14850, USA. Email: zs43@cornell.edu.

1 Introduction

Diffusion processes have proven to be mostly successful in finance over the past three decades in modeling the dynamics of for instance interest rates, stock prices, exchange rates and option prices. On the one hand, the continuous flow of information into financial markets makes it intuitive and necessary to use continuous-time models among which diffusion models may be the most extensively used. On the other hand, the development of stochastic calculus offers us elegant mathematical tools for solving many important problems in finance when diffusion models are used. However, while economic theories have implications about 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. As a result, a great number of parametric diffusion models have been proposed in the literature, see for example Ait-Sahalia(1996a), Ahn and Gao(1999), Chan, Karolyi, Longstaff and Sanders(1992), Cox, Ingersoll and Ross(1985), and Vasicek(1977).

Generally a parametric specification of a diffusion model will essentially specify the whole dynamics of the underlying process, for example, the transition density or the infinitesimal operator of it as a Markov process. Therefore, model misspecification may yield misleading conclusions about the dynamics of the process by rendering inconsistent parameter estimators and their variance-covariance matrix estimators. In practice, such a mis-specified model may result in large errors in pricing, hedging and risk management. Therefore, the development of reliable specification tests for diffusion models is necessary to tackle such problems. However, although in the past decade or so substantial progress has been made in developing estimation methods[1], both parametrically and non or semi-parametrically(see Ait-Sahalia(2002b), Jiang and Knight(1997), Kristensen (2008a,b), Stanton(1997). Bandi and Phillips(2003), for example), relatively little effort has been devoted to specification and evaluation of diffusion models.

In this study, we will develop an omnibus test for the specification of diffusion models based on the infinitesimal operator which is an alternative characterization of the whole dynamics of the process to transition function or transition density used by Ait-Sahalia, Fan and Peng(2008), Chen and Hong(2008a), Corradi and Swanson (2005), and Hong and Li(2005). By the celebrated "martingale problems" developed by Strook and Varadhan(1969), the identification of the diffusion process is equivalent to a "martingale hypothesis" for the processes which come from the transformation of the original diffusion process implied by the "martingale problems". I then check the "martingale hypothesis" for these processes by extending Hong's(1999) generalized spectral approach to a multivariate generalized spectral derivative based test which is particularly powerful against alternatives with zero autocorrelation but a nonzero conditional mean and which has a convenient one-sided $N(0,1)$ limit distribution[2]. The infinitesimal operator of the diffusion process has the nice property of being a closed-form expression of drift and diffusion terms. This makes my test procedure enjoy many good properties which I will discuss in the following.

Since the specification of a parametric diffusion model usually refers to specifying the so-called drift and diffusion terms, we can roughly summarize the existing researches on specification test of parametric diffusions[3]

[1] Sundaresan (2001) points out that "perhaps the most significant development in the continuous-time field during the last decade has been the innovations in econometric theory and in the estimation techniques for models in continuous time." For other reviews of this literature, see (e.g.) Tauchen (1997).

[2] Hong's(1999) test of martingale hypothesis is in spirit similar to Bierens's(1982) and Bierens and Ploberger's(1997) integrated conditional moment tests for model specification. But the latter two has null limit distributions which are a sum of weighted chi-squared variables with weights depending on the unknown data generating process and cannot be tabulated.

[3] In recent years, there have been some researches which check the generic properties of a continuous time process and which are naturally nonparmatric, including the tests of markov property(Ait-Sahalia, 1997; Chen and Hong 2008b), of jumps(Ait-Sahalia

into two categories: the first is focused on the specification testing of either the drift term or the diffusion term, but not both; the second is concentrated on specification of both drift and diffusion terms which determine the whole dynamics of the process. Corradi and White(1999), Li(2007), Kristensen(2008a,b) Fan and Zhang(2003), and Gao and Casas(2008), for example, belong to the first category. The test I will propose belongs to the second category. Therefore, I will mainly review here the related research in this area and compare my test to theirs.

Ait-Sahalia (1996a) developed probably the first nonparametric test for univariate diffusion models. By observing that the drift and diffusion terms completely characterize the stationary (or marginal) density of a diffusion model, Ait-Sahalia (1996a) checks the adequacy of the diffusion model by comparing the model-implied stationary density with a smoothed kernel density estimator based on discretely sampled data. Hong and Li (2005) proposed an omnibus nonparametric specification test based on the transition density, which depicts the full dynamics of a diffusion process. Their idea is that when a diffusion model is correctly specified, the probability integral transform of data via the model-implied transition density is $i.i.d\ U[0,1]$. Hong and Li (2005) then check the joint hypothesis of $i.i.d\ U[0,1]$ by using a smoothed kernel estimator of the joint density of the probability integral transform series. As by-products of the Efficient Method of Moments(EMM) algorithm, a χ^2 test for model misspecification and a class of appealing diagnostic t-tests that can be used to gauge possible sources for model failure were proposed in Gallant and Tauchen (1996). These test can be used to test continuous-time models. Their idea is to match the model-implied moments to the moments implied by a seminonparametric (SNP) transition density for observed data.

Many other tests have appeared recently for univariate diffusion models based on the transition density directly. Both Ait-Sahalia, Fan and Peng (2008) and Chen, Gao and Tang (2008) proposed some tests by comparing the model-implied parametric transition density and distribution function with their nonparametric counterparts with latter using a nonparametric empirical likelihood approach. Corradi and Swanson (2005) introduced two bootstrap specification tests for diffusion processes. The first, for one-dimensional case, is a Kolomogorov type test based on comparison of the empirical cumulative distribution function(CDF) and the model-implied parametric CDF. The second, for multidimensional or multifactor models characterized by stochastic volatility, compares the empirical distribution of the actual data and the empirical distribution of the (model) simulated data. Noticing most of the tests for diffusions only apply for the univariate case, Chen and Hong(2008a) considered a test for multivariate continuous time models based on the conditional characteristic function(CCF) which is the Fourier transform of the transition density. Given the equivalence between the transition density and the CCF, they identify the correct specification of the process as a martingale difference sequence characterization and then generalize Hong's(1999) generalized spectral approach to propose a test.

Compared to all the tests introduced above, my approach has several advantages. First, like the tests based on the transition density which characterize the whole dynamics of the process, my test significantly improves the size and power performance of the marginal density-based test thanks to the use of the infinitesimal generator and the transform based on martingale problems. Although Ait-Sahalia's(1996a) marginal density-based test is convenient to implement, it may easily pass over a misspecified model that has a correct stationary density. In contrast, my infinitesimal operator and martingale problem based test can pick them up effectively because infinitesimal operator is also able to identify the whole dynamics of the process and the martingale

2002a; Ait-Sahalia and Jacod 2008; Lee and Mykland 2008; Fan and Fan 2008), of generic diffusion hypothesis(Kanaya 2007), and so on.

problem is an equivalent characterization of diffusion process. Therefore, my test is omnibus, unlike Gallant and Tauchen's(1996) EMM tests which, as Tauchen (1997) points out, are not consistent against any model misspecification because they are based on a seminonparametric score function rather than the transition density itself.

Second, my procedure can be used to check both the univariate and multivariate models in a convenient way. In fact the transformation based on the martingale problems simplify the test of the specification of a multivariate diffusion process to the test of the martingale hypothesis for many univariate processes. Moreover, the test constructed via the generalized spectral derivative has a convenient one-sided $N(0,1)$ limit distribution and this further simplifies the testing of multivariate diffusion process. In contrast, as discussed in Chen and Hong(2008a), many tests for univariate models introduced above cannot or at least are difficult to be extended to test the multivariate models. For example, Hong and Li's (2005) approach cannot be extended to a multivariate context. The reason is that the probability integral transform of data with respect to a model-implied multivariate transition density is no longer $i.i.d\ U[0,1]$, even if the model is correctly specified. Although Hong and Li (2005) evaluate multivariate affine term structure models for interest rates by using the probability integral transform for each state variable, it may fail to detect misspecification in the joint dynamics of state variables. In particular, their test may easily overlook misspecification in the conditional correlations between state variables. Chen and Hong(2008a) do have the ability to check multivariate diffusion models but their test depends crucially on the availability of closed-form CCF. There will be additional computation burden if CCF does not have a closed-form and numerical methods have to be employed to obtain the CCF from transition density.

Third, my infinitesimal operator based procedure requires nothing except the drift and diffusion terms and hence is simple to implement. This is due to the fact that the infinitesimal operator has always an explicit closed-form expression in terms of the drift and diffusion functions. It is well known that the transition density of most continuous time models has no closed form. As a result, some techniques to approximate the transition density is required in the transition based tests(see Hong and Li (2005), Ait-Sahalia, Fan and Peng (2008)), for example, the simulation methods of Pedersen(1995) and Brandt and Santa-Clara(2002), the Hermite expansion approach of Ait-Sahalia (2002b), or for affine diffusions, the closed-form approximation of Duffie, Pedersen, and Singleton(2003) and the empirical characteristic function approach of Singleton (2001) and Jiang and Knight(2002). Although the asymptotic distribution of some tests(like Hong and Li(2005)) is not affected by the estimation uncertainty, the use of the transition density may not be computationally convenient and may affect the finite-sample performance of the test. In contrast, the infinitesimal operator always has an explicit closed-form expression which can be identified by the drift and diffusion terms. Therefore, we do not need any approximation technique and the test is easy to implement and computationally convenient.

Fourth, the infinitesimal operator based martingale characterization of diffusion process can reveal separate information about the specification of drift and diffusion terms or even their interactions. This is a property which no other approaches enjoy so far. Although other methods can truly check the specification of the drift or diffusion terms by nonparametrically smoothing only one of them, the infinitesimal operator based martingale characterization proposed in this study brings up this type of information in an essential way. In fact, the transformed martingale processes based on infinitesimal operator and "martingale problems" contain separate information about the drift and diffusion terms or their interactions. This motivates me to suggest several feasible tests which are to do separate inference to explore the sources when rejection of a parametric form happens.

Of course, this is not the first time the operator methods are used for econometric inference of continuous time processes. See Ait-Sahalia, Hansen and Scheinkman(2004) for a survey. Hansen and Scheinkman(1995) use infinitesimal operator based moment conditions to derive a GMM-type estimator for time reversible markov processes, including diffusion models. However, the moment conditions used there can only identify the diffusion process up to scale and therefore are not a complete characterization. Moreover, it has to assume the reversibility. Kessler and Sorenson(1996) suggest an estimating equation estimator using the conditional moment restrictions implied by knowledge of an eigenfunction of the infinitesimal operator. But their method has an essential impediment, i.e., it is difficult to compute the implied eigenfunctions from the parametric drift and diffusion terms. Hansen, Scheinkman and Touzi(1998) nonparametrically identify scalar diffusion processes via a conveniently chosen eigenvalue-eigenfunction pair of the conditional expectation operator over a unit interval of time. Hansen and Scheinkman(2003) develop a semi-group theory for Markov pricing with semi-group representing a family of operators that assigns prices today to payoffs that are functions of the Markov state in the future. The relation between the infinitesimal operator and semi-group of operators is that the former is the generator of the latter(see Section 2 in the following for details).

The studies using operator methods above are mainly about identification and estimation. For testing prblems, a new test also based on infinitesimal operator has been proposed recently by Kanaya(2007). This test is mainly to check whether a continuous-time Markov process is a diffusion process instead of checking whether a parametric form of the diffusion process is correct. In other words, the latter assume that the underlying process is a diffusion and check if the parametric specification of the drift and diffusion terms is right, while Kanaya's (2007) test assume the underlying process is only a continuos time Markov process and check if the process is truly a diffusion. It is based on the direct comparison of Nadaraya-Watson type estimators(evaluated at some test functions) for the general continuous-time Markov process and the diffusion process.

Compared to these econometric studies using operator methods, my infinitesimal operator based martingale approach has several nice advantages over them. First, my martingale characterization is a complete identification of the diffusion process unlike Hansen and Scheinkman's(1995) identification up to scale. Second, the martingale identification implies closed-form expressions in terms of drift and diffusion coefficients while the eigenfunctions used in Kessler and Sorenson(1996) and Hansen, Scheinkman and Touzi(1998) are difficult to obtain in general. Third, my proposed test procedure is a specification testing for the parametric diffusion process while Kanaya(2007) is checking the generic diffusion hypothesis. In fact, to the best of my knowledge, my test procedure is the first specification testing for parametric diffusion processes via operator methods. This is important since operators are an important tool to analyze and characterize continuous time stochastic processes and most econometric researches using operators so far are only for identification and estimation. Fourth, my test procedure can be extended to nonparametrically test the generic diffusion hypothesis checked in Kanaya(2007) while the latter's approach is not convenient to check the parametric specification for which my test is powerful. Therefore, it is a more general procedure based on infinitesimal operator than Kanaya(2007). The idea is to nonparametrically estimate the drift and diffusion terms and then use the similar test procedure to check the martingale property of the transformed processes. Of couse, more work needs to be done relative to that in this paper because the convergence rates of nonparametric estimators are slower than those for parametric estimators which are usually $\sqrt{n}$ for sample size n. This research is being inverstigaed and will be reported soon. Fifth, my procedure can conveniently check the multivariate diffusion models but Kanaya's (2007) test is not easy to extend for this situation. For Kanaya's (2007) test of generic diffusion hypothesis, "it may not necessarily be an easy task as hinted by Rogers and Williams(2000, p.243):'The moral is that for dimension $n \geq 2$, infinitesimal operators are not really the right things to look at'"(Kanaya 2007).

The paper is organized as follows. Section 2 clarifies the relationship between a stochastic differential equation and a diffusion process and then introduces the hypotheses of interest for a multivariate time-homogeneous diffusion. Section 3 discusses the construction of the test by a multivariate generalized spectral derivative approach. Test procedures for doing separate inference are suggested in Section4. Asymptotics of the test are presented in Section 5, including the asymptotic distribution and asymptotic power. The applicability of a data-driven lag order is also justified and a plug-in method considered. In Section 6, I examine the finite sample performance of the test procedure by Monte Carlo simulations. Section 7 concludes and discusses possible future researches around the infinitesimal operator based martingale characterization. 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

As we know, the diffusion models in finance are usually specified in terms of a stochastic differential equation:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{2.1}$$

where W_t is a $d \times 1$ standard Brownian motion in $\mathbb{R}^d$, $b : E \subset \mathbb{R}^d \to \mathbb{R}^d$ is a drift function(i.e., instantaneous conditional mean) and $\sigma : E \to \mathbb{R}^{d\times d}$ is a diffusion function(i.e., the instantaneous condition standard deviation). We will call (2.1) a SDE-diffusion process. Obviously, $\{X_t\}$ is a multivariate process.

What we are interested in is to test a parametric form or specification of a SDE-diffusion, i.e., the true drift

$$b^0 \in \mathcal{M}_b \triangleq \{b(\cdot, \theta), \theta \in \Theta\}$$

and the true diffusion

$$\sigma^0 \in \mathcal{M}_\sigma \triangleq \{\sigma(\cdot, \theta), \theta \in \Theta\} \tag{2.2}$$

where Θ is a finite-dimensional parameter space. We say that the model $\mathcal{M}_b$ and $\mathcal{M}_\sigma$ are correctly specified for drift $b^0(X_t)$ and diffusion $\sigma^0(X_t)$ respectively, if

$$H_0 : P[b(X_t, \theta_0) = b^0(X_t),\, \sigma(X_t, \theta_0) = \sigma^0(X_t)] = 1, \text{ for some } \theta_0 \in \Theta \tag{2.3}$$

The alternative hypothesis is that there exists no parameter value $\theta \in \Theta$ such that $b(X_t, \theta)$ and $\sigma(X_t, \theta)$ coincide with $b^0(X_t)$ and $\sigma^0(X_t)$ respectively, that is

$$H_A : P[b(X_t, \theta) = b^0(X_t),\, \sigma(X_t, \theta) = \sigma^0(X_t)] < 1, \text{ for all } \theta \in \Theta \tag{2.4}$$

We will test whether a continuous time SDE-diffusion is correctly specified using $\{X_{\tau\Delta}\}_{\tau=1}^n$, a discrete sample of $\{X_t\}$ observed over a time span T at interval Δ, with sample size $n = T/\Delta$.

Since in this study I am relying on an alternative characterization of a continuous-time Markov process to transition density, i.e., the infinitesimal operator, while in finance a diffusion process is usually specified via a stochastic differential equation(SDE), I will discuss frist some related mathematical concepts and clarify their relationship. By Rogers and Williams(2000, Ch III.1), a continuous time Markov process is defined as follows:

Definition1: A Markov process $X = (\Omega, \{\mathcal{F}_t\}, \{X_t\}, \{P_t\}, \{P^x, x \in E\})_{t \geqq 0}$ with state space (E, ε) is an E-valued stochastic process adapted to the sequence of σ-algebras $\{\mathcal{F}_t\}$ such that

for $0 \leqq s \leqq t$ and $x \in E$, $E^x[f(X_{s+t})|\mathcal{F}_s] = (P_t f)(X_s)$, P^x-*a.s.*

where $\{P_t\}$ is a transition function on (E, ε), i.e., a family of kernels $P_t : E \times \varepsilon \to [0, 1]$ such that

(i): for $t \geqq 0$ and $x \in E$, $P_t(x, \cdot)$ is a measure on ε with $P_t(x, E) \leqq 1$

(ii): for $t \geqq 0$ and $\Gamma \in \varepsilon$, $P_t(\cdot, \Gamma)$ is ε-measurable

(iii): for $s, t \geqq 0$, $x \in E$ and $\Gamma \in \varepsilon$,

$$P_{t+s}(x, \Gamma) = \int_E P_s(x, dy) P_t(y, \Gamma) \tag{2.5}$$

In this definition, the Markov property is characterized by the transition function (or transition density when the density of transition function exists) and (2.5) is the so-called Chapman-Kolmogrov equation. An alternative and equivalent characterization is the indued famliy $\{P_t\}$ which is a set of positive bounded operators with norm less than or equal to 1 on $b\varepsilon$(bounded and ε-measurable functions) and which is defined by:

$$P_t f(x) \equiv (P_t f)(x) = \int_E P_t(x, dy) f(y) \tag{2.6}$$

In this case, the markov property is expressed as the following semi-group property equivalent to the Chapman-Kolmogrov equation:

$$P_s P_t = P_{s+t}, \text{for any } s, t \geqq 0 \tag{2.7}$$

Both transition function and the semi-group of operators characterize the Markov process and interact with the sample-path property of the process. However, since the general Markov process consists of too many processes and is too broad, we choose to focus on the more interesting subclass, Feller process. By Rogers and Williams(2000, Ch III.6), Feller process is defined as follows :

Definition2: The transition function $\{P_t\}_{t \geqq 0}$ of a Markov process is called a Feller transition function if

(i): $P_t C_0 \subset C_0$ for all $t \geqq 0$

(ii): for any $f \in C_0$ and $x \in E$, $P_t f(x) \to f(x)$ as $t \downarrow 0$, where $C_0 = C_0(E)$ is the space of real-valued, continuous functions on E which vanish at infinity and C_0 is endowed with the sup-norm.

Feller process has good path properties[4] and is also general enough to contain most processes we are interested in, for example, Feller diffusion which will be defined below and has been extensively used in finance,

[4] 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

and Levy process including Poisson process and Compound process which has received more and more attention in finance recently(see Schoutens, 2003). For Feller processes, we will consider another characterization, the infinitesimal operator, other than the transition function and semi-group of operators introduced above which are for the general Markov process.

Definition3: A function $f \in C_0$ is said to belong to the domain $D(A)$ of the infinitesimal operator of a Feller process X if the limit

$$Af = \lim_{t \downarrow 0} \frac{P_t f - f}{t} \tag{2.8}$$

exists in C_0. The linear transformation $A : D(A) \to C_0$ is called the infinitesimal operator of the process.

Immediately from the Definition3, we see that for $f \in D(A)$, it holds P-$a.s.$ that

$$E\left(\frac{f(X_{t+h}) - f(X_t)}{h} | \mathcal{F}_t\right) = Af(X_t) + o(h), \text{ as } h \downarrow 0 \tag{2.9}$$

In this sense, the infinitesimal operator indeed describes the movement of the process in an infinitesimally small amount of time. Therefore, intuitively the infinitesimal operator characterizes the whole dynamics of a Feller process because the time is continuous here[5].

So far we have had Feller process for which three complete characterization of the dynamics are available: transition function(or transition density), semi-group of operators and infinitesimal operator. In the following, we will consider probably the most popular processes used in continuous time finance which we are going to test, the diffusion processes. By Rogers and Williams(2000, ChIII.13),

Definition4: A Feller process with state space $E \subset \mathbb{R}^d$ is called a Feller diffusion if it has continuous sample paths and the domain of its infinitesimal operator contains the function space $C_c^\infty(int(E))$ which is the space of infinitely differentialbe functions with compact support contained in the interior of the state space E.

We can see that the Feller diffusion is defined through the combination of the sample path properties and the restrictions imposed on the infinitesimal operator. A very convenient property of Feller diffusion is that its infinitesimal operator has an explict form. According to Kallenberg(2002, Thm 19.24) and Rogers and Williams (2000, Vol1, Thm III.13.3 and Vol2, Ch V.2), for a Feller diffusion $\{X_t\}$, there exist some functions $a_{i,j}$ and $b_i \in C(\mathbb{R}^d)$ for $i, j = 1, \cdots, d$ where $(a_{i,j})_{i,j=1}^d$ forms a symmetric nonegative definite matrix such that the infinitesimal operator is

$$Af(x) = \sum_{i=1}^{d} b_i(x) f_i^{'}(x) + \frac{1}{2} \sum_{i,j=1}^{d} a_{i,j}(x) f_{i,j}^{''}(x) \tag{2.10}$$

for $f \in D(A)$ and $x \in \mathbb{R}^d$.

Now we have arrived at the Feller diffusion and its infinitesimal opetator which has a closed form. Also we

[5] Rigorouly, 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)). Therefore, infinitesimal operator does determine the whole dynamics of the process.

have SDE-diffusion (2.1) at hand. Then what is the relationship between them? By Rogers and Williams(2000, ChV.2 and V.22), under some regularity conditons, they are equivalent. That is, for a Feller diffusion as in Definition4, there is a corresponding SDE-diffusion and also a SDE-diffusion like (2.1) is a Feller diffusion, where the function $b(\cdot)$ are the same and $a = \sigma\sigma^T$, i.e., $a_{ij}(x) = \sum_{k=1}^{d} \sigma_{i,k}(x)\sigma_{j,k}(x)$. Therefore, the SDE-diffusion which has been analyzed extensively in continuous-time finance and which belongs to the class of Feller process also has (2.10) as the closed-form infinitesimal operator.

Since many processes in finance are univariate(for example, see Ait-Sahalia 1996a for some univariate models for term structure of interest rate), I consider a univariate diffusion here for illustration. A univariate diffusion is defined as $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ with W_t a 1-dimensional standard Brownian motion in $\mathbb{R}$, $b : E \subset \mathbb{R} \to \mathbb{R}$ a drift function and $\sigma : E \to \mathbb{R}$ a diffusion function. Then by (2.10) and the discusstion above, the infinitesimal operator for this univariate diffusion is

$$Af(x) = b(x)f^{'}(x) + \frac{1}{2}\sigma^2(x)f^{''}(x) \tag{2.11}$$

Clearly the first term involving the first derivative of function $f(\cdot)$ is related to the dynamics of drift and the second term involving the second derivative of function $f(\cdot)$ to the dynamics of diffusion function. This is consistent with the intuition that drift describes the dynamics of mean and the diffusion describes that of variance of the process(see Nelson 1990 for more discussion which proves that the diffusion process is the approximation of an ARCH process). However, this intuition is not always right due to the continuous nature of the time. Consider the infinitesimal changes of this univariate diffusion process. By (2.9) and (2.11), for any $f \in D(A)$, it holds P-$a.s.$ that

$$E\left(\frac{f(X_{t+h}) - f(X_t)}{h}|\mathcal{F}_t\right) = b(X_t)f^{'}(X_t) + \frac{1}{2}\sigma^2(X_t)f^{''}(X_t) + o(h), \text{ as } h \downarrow 0 \tag{2.12}$$

Therefore, the dynamics of $\{X_t\}$ are characterized completely by the drift and diffusion coefficients, including the conditional probability law. However, in discrete time series models, the mean and variance solely cannnot determine the complete conditional probability law unless it is gaussian. Therefore, it is not right to simply think of drift and diffusion terms as the continuous time counterparts of conditional mean and variance respectively. In fact, the conditional mean of the process $\{X_t\}$, $E[X_{t+h}|X_t]$ for a fixed $h > 0$ is a function of both the drift $b(\cdot)$ and diffusion $\sigma(\cdot)$ instead of the drift solely(see Ait-Sahalia 1996a).

From the discussions above, for SDE-diffusion which is also a Feller diffusion, there are at least two characterizations we can use to identify the whole dynamics of the process: the transition function and the the closed-form infinitesimal operator which is also the generator of the third characterization, semi-group of operators. The former, also well known as transition density when the density of transition function exists, has been the primary tool to analyze the diffusion process, not only in estimation(see Ait-Sahalia 2002b) but also in the construction of specification tests (see Hong and Li 2005, Ait-Sahalia, Fan and Peng 2008). However, as we discussed in Secion1, specification of drift and diffusion terms rarely give a closed-form transition density. In contrast, (2.10) and (2.11) tell us that the infinitesimal operator does have a direct and explicit expression and this nice property makes it a convenient tool for analyzing the diffusion process. It has already been used in identification and estimation problems as discussed above and the idea of constructing a specification test for diffusion process comes up naturally.

To construct a test of diffusion based on infinitesimal operator, I 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 check. Let me first define the martingale problem(see, Karatzas and Shreve(1991), Ch5.4):

Definition5: A probability measure P on $\left(C[0,\infty)^d, \mathcal{B}(C[0,\infty)^d)\right)$ under which

$$M_t^f = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}f)(X_s)ds \text{ is a martingale for every } f \in D(\mathcal{A}) \quad (2.13)$$

is called a solution to the martingale problem associated with the operator $\mathcal{A}$.

How is the martingale problem related to the SDE-diffusion? As we know, 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). Intuively, 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 for general drift and diffusion terms, a strong solution may not exist;in this case, probablists 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 in fact equivalent to the weak solution of a SDE as shown by the following:

Theorem1: The process $\{X_t\}$ is a weak solution to the SDE (2.1) if and only if it satisfies the martingale problem of Definition5 with A as the infinitesimal operator of $\{X_t\}$ as defined in (2.10).

Now we have shown that the weak solution of a SDE is equivalent to the martingale problem. When strong solution exists the weak solution will coincide with it. Hence it is enough to consider the weak solution identification for doing econometric inference because regularity conditions for the existence of strong solution are usually satisfied and thus imposed in analysis[6](see Protter 2005 for some regularity Lipschitz conditions for the existence and uniqueness of a strong solution to a SDE).

By Theorem1 and (2.13), the correct specification of a SDE-diffusion is equivalent to whether the martingale problem is satisfied, implying that the hypotheses of interest H_0 in (2.3) versus H_A in (2.4) can be equivalently written as:

H_0 : For some $\theta_0 \in \Theta$, $M_t^f(\theta_0) = f(X_t) - f(X_0) - \int_0^t (A_{\theta_0} f)(X_s)ds$ is a martingale for every $f \in D(A)$,

$$\text{where } A_{\theta_0} f(x) = \sum_{i=1}^d b_i(x;\theta_0) f_i^{'}(x) + \frac{1}{2}\sum_{i,j=1}^d a_{i,j}(x;\theta_0) f_{i,j}^{''}(x) \text{ and } a_{ij}(x;\theta_0) = \sum_{k=1}^d \sigma_{i,k}(x;\theta_0)\sigma_{j,k}(x;\theta_0) \quad (2.14)$$

Versus

H_A : For all $\theta \in \Theta$, $M_t^f(\theta) = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}_\theta f)(X_s)ds$ is not a martingale for some $f \in D(\mathcal{A})$,

[6] I thank Professor Philip Protter for suggesting this point to me.

$$\text{where } \mathcal{A}_\theta f(x) = \sum_{i=1}^{d} b_i(x;\theta) f_i^{'}(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x;\theta) f_{i,j}^{''}(x) \text{ and } a_{ij}(x;\theta) = \sum_{k=1}^{d}\sigma_{i,k}(x;\theta)\sigma_{j,k}(x;\theta) \qquad (2.15)$$

Now we have transformed the correct specification hypothesis of a multivariate time-homogeneous diffusion into a martingale hypothesis for some new processes based on the infinitesimal operator and martingale problems which is very convenient to check. Observe from (2.14) that what we have to is only to check the martingale property for the transformed processes M_t^f for every $f \in D(\mathcal{A})$. However, there are usually an infinite number of functions $f(\cdot)$ in the domain $D(\mathcal{A})$ which are usually called test functions(note that $D(\mathcal{A})$ contains the function space $C_c^\infty(int(E))$ as a subset for Feller diffusion defined in Definition4). Hence we unfortunately have to check the martingale property for infinitly many processes $\{M_t^f\}$ for test function $f \in D(\mathcal{A})$. It is definitely impossible in practice and we need a subclass of $D(\mathcal{A})$ which not only consists of finitely many function forms but also plays the same role as $D(\mathcal{A})$ does. Luckily, the following celebrated theorem gives an equivalent subclass by which a practical test procedure can be constructed easily[7].

Theorem2: The process $\{X_t\}$ is a weak solution to the SDE in (2.1) if it satisfies the martingale problem of Definition5 with $\mathcal{A}$ as the infinitesimal operator of $\{X_t\}$ for the choices $f(x) = x_i$ and $f(x) = x_i x_j$ with $1 \leq i, j \leq d$.

At first glance, this result may appear confusing because $f(x) = x_i$ and $f(x) = x_i x_j$ do not belong to $D(A)$ which is a subset of $C_0(\mathbb{R}^d)$. To get an intuition for this important result, let me choose sequences $\{g_i^{(K)}\}_{K=1}^\infty$ and $\{g_{ij}^{(K)}\}_{K=1}^\infty$ in function space $C_0(\mathbb{R}^d)$ such that $g_i^{(K)}(x) = x_i$ and $g_{ij}^{(K)} = x_i x_j$ for $\|x\| \leq K$. If $M^{g_i^{(K)}}$ and $M^{g_{ij}^{(K)}}$ are martingales, then M^{x_i} and $M^{x_{ij}}$ are local martingales. A similar result to Theorem1 with local martingale replacing martingale then tells us that $\{X_t\}$ is a weak solution to the SDE in (2.1). See the proof in Appendix for details. Of course, the converse of Theorem2 only holds with local martingale replacing martingale. However, since examples which are local martingales but not martingales are few and too artificial in certain sense even when they exist[8], I regard them as almost the same and do not pay much attention to their difference in this study[9]. Actually, by imposing certain regularity conditions, a local martingale can become a martingale(see Protter 2005 for such techniqual conditions). But I do not explore them here since it is not the focus of this study and could certainly distract the attention. To sum up, Theorem2 implies that the hypotheses of interest H_0 in (2.14) versus H_A in (2.15) can be equivalently written as:

H_0 : For some $\theta_0 \in \Theta$

[7] We can also reduce the space of test functions to an equivalent subclass by the method considered in Kanaya(2007) which is based on the concept of a core and "approximation" theory. Since my reduced space of test functions constructed by Theorem2 is much more simple and intuitive than that in Kanaya(2007), I do not use that method here. Also see Hansen and Scheinkman(1995) and Conley, Hansen, Luttmer and Scheinkman(1997) for more discussions about choices of test functions.

[8] See Karatzas and Shreve(1991), p.168 and 200-201 for some examples which are local martingales but not martingales.

[9] When the difference really matters, the local martingale property can be used in the specification testing of diffusion models. The idea is to use the fact that the time-changed continuous local martingale by quadratic variation is a standard Brownian Motion(see Andersen, Bollerslev & Dobrev(2007) and Park(2008) for details). Since this approach is closely related to time-dependent diffusion models and the test procedure will be very different, I do not pursue it here. But the research on it is being investigated and will be reported soon.

$$M_t^{x_i}(\theta_0) = X_t^i - X_0^i - \int_0^t b_i(X_s;\theta_0)ds$$
$$M_t^{x_i,x_j}(\theta_0) = X_t^{x_i,x_j} - X_0^{x_i,x_j} - \int_0^t \left[b_i(X_s;\theta_0)X_s^j + b_j(X_s;\theta_0)X_s^i + \sum_{k=1}^d \sigma_{i,k}(X_s;\theta_0)\sigma_{j,k}(X_s;\theta_0)\right] ds \tag{2.16}$$

are martingales for $1 \leq i, j \leq d$.

Versus

H_A : For all $\theta \in \Theta$

$$\text{either } M_t^{x_i}(\theta) \text{ or } M_t^{x_i,x_j}(\theta) \text{ is not a martingale for some } i, j = 1, \cdots, d \tag{2.17}$$

where $M_t^{x_i}(\theta)$ and $M_t^{x_i,x_j}(\theta)$ are defined as in (2.16) with θ replacing θ_0

This greatly simplifes the hypothesis and makes the testing of the specification completely practical. Two things are worth pointing out here. The first one is that my hypothesis of correct specification can be expressed explicitly by the drift and diffusion terms. Therefore, any spcification of the diffusion model can be tested directly without computation of transition density and the asymptotic distribution is completely free of estimation uncertainty as long as the estimator is $\sqrt{n}$-consistent. In contrast, the transition density based methods like Hong and Li(2005) or Ait-Salalia, Fan and Peng(2008) have to approximate the model-implied transition density first because the transition density hardly has a closed-form. Furthermore, this explicit expression enables my test procedure to be free of the so-called "nuisance parameter" problem encountered in Chen and Hong(2008a) which constructs a test of multivariate diffusion based on conditional characteristic function with nuisance parameters. The second is that my procedure transforms the specification test of a multivariate d-dimensional diffusion process into tests of martingale property for $d' = (d^2 + 3d)/2$ univariate processes. This makes my test procedure particularly covenient for the specification of multivariate diffusion processes which is very difficult by the transition density based methods.

Similarly for the univariate diffusion process $\{X_t\}$ by (2.11), the correct specification hypotheses of interest are:

H_0 : For some $\theta_0 \in \Theta$

$$M_t^x(\theta_0) = X_t - X_0 - \int_0^t b(X_s;\theta_0)ds$$
$$M_t^{x^2}(\theta_0) = X_t^2 - X_0^2 - \int_0^t \left[2b(X_s;\theta_0)X_s + \sigma^2(X_s;\theta_0)\right] ds \text{ are both martingales} \tag{2.18}$$

versus

H_A : For all $\theta \in \Theta$

$$\text{either } M_t^x(\theta) \text{ or } M_t^{x^2}(\theta) \text{ is not a martingale}$$
$$\text{where } M_t^x(\theta) \text{ and } M_t^{x^2}(\theta) \text{ are defined as in (2.18) with } \theta \text{ replacing } \theta_0 \tag{2.19}$$

For the convenience of constructing a test procedure, I further state the following equivalent hypotheses of correcet specification in terms of the $m.d.s.$property for the tranformed processes.

H_0 : For some $\theta_0 \in \Theta$, $E\left[Z_t(\theta_0)|\mathcal{I}_{t'}\right] = 0$ for any $t^{'} < t$, where $\mathrm{I}_{t'} = \sigma\{X_{t"}\}_{t"<t'}$ is the sigma-field generated by the past information of $\{X_t\}$ at time $t^{'}$ and $Z_t(\theta_0)$ is a vector with components for $i, j = 1, \cdots, d$

$$\begin{aligned} Z_t^i(\theta_0) &= M_t^{x_i}(\theta_0) - M_{t-1}^{x_i}(\theta_0) = X_t^i - X_{t-1}^i - \int_{t-1}^t b_i(X_s;\theta_0)ds \\ Z_t^{i,j}(\theta_0) &= M_t^{x_i x_j}(\theta_0) - M_{t-1}^{x_i x_j}(\theta_0) \\ &= X_t^i X_t^j - X_{t-1}^i X_{t-1}^j - \int_{t-1}^t \left[b_i(X_s;\theta_0)X_s^j + b_j(X_s;\theta_0)X_s^i + \sum_{k=1}^d \sigma_{i,k}(X_s;\theta_0)\sigma_{j,k}(X_s;\theta_0)\right] ds \quad (2.20) \end{aligned}$$

versus

$$\begin{aligned} H_A \quad : \quad & \text{For all } \theta \in \Theta,\ E\left[Z_t(\theta)|\mathcal{I}_{t'}\right] \neq 0 \text{ for any } t^{'} < t, \\ & \text{where } \mathcal{I}_{t'} \text{ and } Z_t(\theta) \text{ is defined as in (2.20) with } \theta \text{ replacing } \theta_0 \quad (2.21) \end{aligned}$$

Corresponding to (2.20) and (2.21) for multivariate diffusion models, the $m.d.s.$representation of hypotheses of intersest for univariate case is:

H_0 : For some $\theta_0 \in \Theta$, $E\left[Z_t(\theta_0)|\mathcal{I}_{t'}\right] = 0$ for any $t' < t$, where $Z_t(\theta_0) = \left(Z_t^x(\theta_0), Z_t^{x^2}(\theta_0)\right)'$, $\mathcal{I}_{t'}$ is defined as in (2.20), and

$$\begin{aligned} Z_t^x(\theta_0) &= M_t^x(\theta_0) - M_{t-1}^x(\theta_0) = X_t - X_{t-1} - \int_{t-1}^t b(X_s;\theta_0)ds \\ Z_t^{x^2}(\theta_0) &= M_t^{x^2}(\theta_0) - M_{t-1}^{x^2}(\theta_0) = X_t^2 - X_{t-1}^2 - \int_{t-1}^t \left[2b(X_s;\theta_0)X_s + \sigma^2(X_s;\theta_0)\right] ds \quad (2.22) \end{aligned}$$

versus

$$\begin{aligned} & H_A\text{: For all } \theta \in \Theta,\ E\left[Z_t(\theta)|\mathcal{I}_{t'}\right] \neq 0 \text{ for any } t' < t, \\ & \quad \text{where } \mathcal{I}_{t'} \text{ is defined as in (2.21) and } Z_t(\theta) \text{ is defined as in (2.22) with } \theta \text{ replacing } \theta_0 \quad (2.23) \end{aligned}$$

3 Test Procedure based on Multivariate Generalized Spectral Derivative

In this section, I shall construct a test procedure of the correct specification hypotheses H_0 versus H_A in (2.20) and (2.21) for the multivariate diffusion process. As an illustration, I shall also present the test procedure for

H_0 versus H_A in (2.22) and (2.23) for univariate diffusion process which is a special case of that for multivariate case. The sample data is discrete in time, i.e., $\{X_{\tau\Delta}\}_{\tau=1}^n$ observed over a time span T with sampling interval Δ and sample size $n = T/\Delta$. Therefore, the process is in continuous time but the data sample is discrete. This is a general problem in continuous-time series econometrics not only for testing but also for estimation(see Lo (1988) and Ait-Sahalia(1996a,b) for discussions about the estimation of the discretized version of a continuous-time model). Like Ait-Sahalia(1996b), Hong and Li(2005) or Kanaya(2007), I will consider the discrete time implications of the *m.d.s.* property which is derived in continuous time[10]. The asymptotic scheme I use here is $n = T/\Delta \to \infty$. It can be obtained by either infill($\Delta \to 0$) or long span ($T \to \infty$)[11] instead of both and this implies that my test procedure can be applied to both high-frequency and low-frequency data. In contrast, many other papers like Stanton (1997), Bandi and Philips(2003), and Kanaya(2007) assume $\Delta \to 0$ and hence can only be used for high-frequency data.

The null hypothesis is that $E\left[Z_t(\theta_0)|\mathcal{I}_{t'}\right] = 0$ for any $t' < t$, where $\mathrm{I}_{t'}=\sigma\{X_{t''}\}_{t''<t'}$ is the sigma-field generated by the past information of $\{X_t\}$ and $Z_t(\theta_0)$ is a vector with components defined in (2.20). Also by (2.20), this implies

$$E\left[Z_t(\theta_0)|\mathcal{I}_{t'}^Z\right] = 0 \text{ for any } t' < t, \text{ where } \mathcal{I}_{t'}^Z = \sigma\{Z_{t''}(\theta_0)\}_{t''<t'} \tag{3.1}$$

where $\mathcal{I}_{t'}^Z$ is the sigma-field generated by past information of $\{Z_t(\theta_0)\}$[12]. Since the sample data we have is $\{X_t, t = \tau\Delta\}_{\tau=0}^n$ with $n = T/\Delta$, an application of the Law of Iterated expectation as well as (3.1) implies that

$$E\left[Z_{\tau\Delta}(\theta_0)|\mathcal{I}_{\tau-1}^Z\right] = 0, \text{ where } \mathcal{I}_{\tau-1}^Z = \sigma\{Z_{(\tau-1)\Delta}(\theta_0), Z_{(\tau-2)\Delta}(\theta_0), \cdots, Z_\Delta(\theta_0), Z_0(\theta_0)\} \tag{3.2}$$

Observe that (3.2) is a *m.d.s.* property for discrete time process $\{Z_{\tau\Delta}(\theta_0)\}_{\tau=0}^n$ and it is derived as an implication of the *m.d.s.*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) which deal with estimation problems and therefore is free of the discretization errors which are discussed in Lo(1988) in the context of estimation. Moreover, this property is also the reason why my test procedure based on (3.2) is only assuming $n = T/\Delta \to \infty$ for asymptotic theory and applicable to both low and high fiquency data. In contrast, other procedures using the discretization scheme of a continuous time model only apply to high frequency data and henthfore a little restricted(for example, see Gao and Casas(2008) and Fan and Zhang(2003)).

As discussed above, my test procedure will be based on checking whether (3.2) is true or not. However, it is not a trivial task to check this. First, the conditioning information set $\mathcal{I}_{\tau-1}^Z$ has an infinite dimension as $\tau \to \infty$ and then there is a "curse of dimensionality" difficulty associated with testing the *m.d.s.*property. Second, $\{Z_{\tau\Delta}(\theta_0)\}$ may display serial dependence in its higher order conditional moments. Any test should be robust to time-varying conditional heteroskedasticity and higer order moments of unknown form in $\{Z_{\tau\Delta}(\theta_0)\}$. To

[10] The discretization can be justified by Zahle(2008) who proves rigourously that the discrete time processes solving the disctete analogue of the martingale problem approximate weakly the solution of the stochastic differential equation under additional assumption on the moments of the increments.

[11] Bandi and Phillips(2003) argued that both the infill and long-span assumptions are needed to estimate continuous-time (diffusion) process fully non-parametrically.

[12] $\mathcal{I}_{t'}$ can still be used here and this actually simplifies the test statistic in $\widehat{M_0}(p)$ (3.9) because $\{X_t\}$ is only d-dimensional while $\{Z_t\}$ is a $2d'$-dimensional process. The test procedure constructed this way can be called a generalized cross-spectral derivative approach. I do not follow this method here mainly to simplify the notations. And because of the close relationship between $\{X_t\}$ and $\{Z_t\}$ which can be seen from (2.20), it will not matter too much for the final result.

check the $m.d.s.$property of $\{Z_{\tau\Delta}(\theta_0)\}$, I extend Hong's(1999) generalized spectral approach to a multivariate generalized spectral derivative method. The idea is similar to Hong and Lee(2005) which considers testing time series conditional mean models with no prior knowledge of possible alternatives. The difference is that here the process I check for $m.d.s.$property is transformed explicitly from the original process while the process Hong and Lee(2005) check is the estimated residuals from a conditional mean model. Furthermore, the process $\{Z_{\tau\Delta}(\theta_0)\}$ is multivariate but that in Hong and Lee(2005) is only univariate. Therefore, the problem here is more complicated and we need to extend the generalized spectral approach to an multivariate one while keeping the property of being free of "curse of dimensionality". Hong and Lee(2004) does extend the generalized spectral approach to an multivariate case for testing multivariate volatility models. However, what they check there is a condition on the volatility martrix of the estimated residuals and thus is different from (3.2) here which is a condition on the mean. The multivariate generalized spectral derivative approach used here is much more simple and intuitive.

Suppose $\{Z_\tau\}$ is a strictly stationary process with marginal characteristic function $\varphi(u) = E(e^{iu'Z_\tau})$ and pairwise joint characteristic function $\varphi_m(u,v) = E(e^{iu'Z_\tau + iv'Z_{\tau-|m|}})$, where $i = \sqrt{-1}$, $u, v \in \mathbb{R}^{d'}$, and $m = 0, \pm1, \cdots$. The basic idea of the generalized spectrum is to consider the spectrum of the transformed series $\{e^{iu'Z_\tau}\}$. It is defined as

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

where ω is the frequency, and $\sigma_m(u,v)$ is the covariance function of the transformed series:

$$\sigma_m(u,v) \equiv cov(e^{iu'Z_\tau}, e^{iv'Z_{\tau-|m|}}), m = 0, \pm1, \cdots \tag{3.3}$$

Note that the function $f(\omega, u, v)$ is a complex-valued scalar function although Z_τ is a $d' \times 1$ vector. It can capture any type of pairwise serial dependence in $\{Z_\tau\}$, i.e., dependence between Z_τ and $Z_{\tau-m}$ 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 existence of any moment of $\{Z_\tau\}$. This is important in economics and finance because it has been argued that the higher order moments of many financial time series may not exist. 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 $m.d.s.$property over many lags in a pairwise manner, avoiding the "curse of dimensionality" difficulty. 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 will capture 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 $\{Z_\tau\}$, $f(\omega, u, v)$ can be differentiated to capure 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

$$\sigma_m^{(1,0)}(0,v) \equiv \frac{\partial}{\partial u}\sigma_m(u,v)\ |_{u=0} = cov(iZ_\tau, e^{iv'Z_{\tau-|m|}}) \tag{3.4}$$

is a $d'\times 1$ vector. The measure $\sigma_m^{(1,0)}(0,v)$ checks whether the autogregression function $E[Z_\tau \mid Z_{\tau-m}]$ 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 $E[Z_\tau \mid Z_{\tau-m}]=0$, $a.s$[13].

It should be noted that the hypothesis of $E\left[Z_\tau(\theta)\mid \mathcal{I}_{\tau-1}\right]=0$ $a.s.$is not exactly the same as the hypothesis of $E[Z_\tau \mid Z_{\tau-m}]=0$ $a.s.$ for all $\tau>0$. 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 for which $E[Z_\tau \mid Z_{\tau-m}]=0$ $a.s.$ for all $\tau>0$ but $E\left[Z_\tau(\theta)\mid \mathcal{I}_{\tau-1}\right]\neq 0$ $a.s.$may be rare in practice and are thus pathological. Even in cases for which the gap does matter, it can be further narrowed down by using the function $E[Z_\tau \mid Z_{\tau-m}, Z_{\tau-l}]$ which may be called the bi-autoregression function of Z_τ at lags (m,l). An equivalent measure is the generalized third order central cumulant function $\sigma_{m,l}^{(1,0)}(0,v)=cov\left[Z_\tau, \exp\left(iv_1'Z_{\tau-m}+iv_2'Z_{\tau-l}\right)\right]$, where $v=(v_1,v_2)\in\mathbb{R}^{d'}\times\mathbb{R}^{d'}$. This is a straightforward extension of generalized bispctrum ananlysis proposed in Hong and Song(2008) for univariate processes.

In the present context, I supress Δ and θ_0 and then let $Z_\tau\equiv Z_{\tau\Delta}(\theta_0)$ for the simplification of notations. Obvisouly, Z_τ cannot be observed. We can first estimate the parameter θ_0 by the random sample $\{X_{\tau\Delta}\}_{\tau=1}^n$ to get a $\sqrt{n}$-consistent estimator. Then the estimated processes $\widehat{Z}_\tau = Z_{\tau\Delta}(\widehat{\theta})$ is obtained. Examples of $\widehat{\theta}$ are approximated transition density based estimator in Ait-Sahalia(2002b), simulated MLE in Pedersen(1995) and so on. Then we can estimate $f^{(0,1,0)}(\omega,0,v)$ for process $\{Z_\tau(\theta)\}$ by the following smoothed kernel estimator:

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

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

$$\widehat{\varphi}_m(u,v)=\frac{1}{n-|m|}\sum_{\tau=|m|+1}^{n}e^{iu'\widehat{Z}_\tau+iv'\widehat{Z}_{\tau-|m|}} \tag{3.5}$$

Here, $p=p(n)$ is a bandwidth, and $k:\mathbb{R}\to[-1,1]$ is a symmetric kernel. Examples of $k(\cdot)$ include Barlett, 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, $\widehat{f}^{(0,1,0)}(\omega,0,v)$ is consistent for $f^{(0,1,0)}(\omega,0,v)$. See Theorem 3 below.

Under H_0, we have $\sigma_m^{(1,0)}(0,v)=0$ 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 ω:

$$f_0^{(0,1,0)}(\omega,0,v)\equiv\frac{1}{2\pi}\sigma_0^{(1,0)}(0,v)=\frac{1}{2\pi}cov\left(iZ_\tau,e^{iv'Z_\tau}\right),\ \omega\in[-\pi,\pi] \text{ and } v\in\mathbb{R}^{d'} \tag{3.6}$$

[13]See Bierens(1982) and Stinchcombe and White(1998) for discussion on related issure in an $i.i.d.$ context.

which can be consistently estimated by

$$\widehat{f}_0^{(0,1,0)}(\omega, 0, v) = \frac{1}{2\pi}\widehat{\sigma}_0^{(1,0)}(0, v),\ \omega \in [-\pi, \pi] \text{ and } v \in \mathbb{R}^{d'} \tag{3.7}$$

The estimators $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ and $\widehat{f}_0^{(0,1,0)}(\omega, 0, v)$ converge to the same limit undr H_0 and generally converge to different limits under H_A. Thus, any significant divergence between them is evidence of the violation of the MDS property and hence of the mis-specification of the process. We can measure the distance between $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ and $\widehat{f}_0^{(0,1,0)}(\omega, 0, v)$ by quadratic form:

$$\widehat{Q} \equiv \int\int_{-\pi}^{\pi} \left\| \widehat{f}^{(0,1,0)}(\omega, 0, v) - \widehat{f}_0^{(0,1,0)}(\omega, 0, v) \right\|^2 d\omega dW(v) = \sum_{m=1}^{n-1} k^2(m/p)(1 - m/n) \int \left\| \widehat{\sigma}_m^{(1,0)}(0, v) \right\|^2 dW(v) \tag{3.8}$$

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.

My proposed omnibus test statistic for correct specification hypothesis is an appropriately standardized version of $\widehat{Q}$,

$$\widehat{M_0}(p) = \left[\sum_{m=1}^{n-1} k^2(m/p)(n - m) \int \left\| \widehat{\sigma}_m^{(1,0)}(0, v) \right\|^2 dW(v) - \widehat{C}_0(p) \right] / \sqrt{\widehat{D}_0(p)}$$

where

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

$$\begin{aligned} \widehat{D}_0(p) &= 2 \sum_{m=1}^{n-2} \sum_{l=1}^{n-2} k^2(m/p) k^2(l/p) \sum_{a=1}^{d'} \sum_{a'=1}^{d'} \int\int \\ &\times \left| \frac{1}{n - \max(m, l)} \sum_{\tau=\max(m,l)+1}^{n} \left[\widehat{Z}_{a\tau} \widehat{Z}_{a'\tau} \right] \widehat{\psi}_{\tau-m}(v) \widehat{\psi}_{\tau-l}^*(u) \right|^2 dW(u) dW(v) \end{aligned} \tag{3.9}$$

and $\widehat{\psi}_\tau(v) = e^{iv'\widehat{Z}_\tau} - n^{-1} \sum_{\tau=1}^{n} e^{iv'\widehat{Z}_\tau}$. Throughout, all unspecified integrals are taken on the support of $W(\cdot)$. The factors $\widehat{C}_0(p)$ and $\widehat{D}_0(p)$ are approximately the mean and the variance of quadratic form $n\widehat{Q}$. The impact of conditional heteroskedasticity and other time-varying higher order conditional moments has already been taken into account. Note that $\widehat{M_0}(p)$ involves d'- and $2d'$-dimensional numerical integrations which can be computationally cumbersome when d' is large. In practice, one may choose a finite number of grid points symmetric about zero or generate a finite number of points drawn from a uniform distribution on $[-1, 1]^{d'}$. The asymptotic theory allows for both discrete and continuous weighting function for $W_0(\cdot)$ which weigh sets symmetric about zero equally. A continuous weighting function for $W_0(\cdot)$ will ensure good power for $\widehat{M_0}(p)$, but there is a trade-off between computational cost and power gains when choosing a discrete or continuous weighting function. One may expect that the power of $\widehat{M}_0(p)$ will be ensured if sufficiently fine grid points are used.

Since we go such a long way before getting the test statistics, let me summarize my omnibus test procedure: (1) Estimate the parametric diffusion model to get a $\sqrt{n}$-consistent estimator $\widehat{\theta}$ for θ_0[14]; (2) Compute the model implied processes $\{Z_\tau(\widehat{\theta})\}$; (3) Compute the test statistic $\widehat{M}_0(p)$ defined in (3.9) ; (4) Compare the value of $\widehat{M}_0(p)$ with the upper-tailed $N(0,1)$ critical value C_α at level α(this follows from Theorem 3 I will derive in the next section). If $\widehat{M}_0(p) > C_\alpha$, then reject H_0 at level α.

4 Seperate Inference

When a model is rejected using the test procedure above which checks jointly the specification of both drift and diffusion terms, it would be interesting to explore possible sources of the rejection. Specifically, is the rejection due to the misspecified form of drift function or the diffusion function? Having this information in hand, one can try other parametric forms of drift, diffusion or both. This is particularly important when economic theory provides little guidance about the specification of the drift and diffusion, which is usually the case in practice.

However, only several papers are available in this respect and most are focused on the specification of diffusion term, like Corradi & White(1999) and Li(2007). Kristensen(2008a) and Gao & Casas(2008) develop specification tests for both the drift and diffution terms but they need to assume the correct specification of the diffusion term a priori and hence are subject to diffusion misspecification. The tests proposed in Kristensen(2008b) and Fan & Zhang(2003) as well as those suggested although not explored in Li(2007) and Bandi & Philips(2007) do have the ability to check the drift term robust to diffusion misspecification. But the gains are achieved by the cost of nonparametrically estimating the diffusion or drift term which has already been challenged seriously by Chapman & Pearson(2000) and Pristker(1998). Moreover, to do the nonparametric estimation of drift or diffusion terms, high frequency data with sampling interval going to zero is needed(see Stanton 1997) which may not be a valid assumption for daily interest rate data. Kristensen(2008b) does not involve nonparametrically smoothing dirft or diffusion term. But he relies on comparison between a semiparametric implied transition density using nonparametric smoothing for marginal desntiy and a nonparametric directly estimated transition density. It is well known that transition density does not have a closed-form in general and hence simulation methods are used in Kristensen(2008b). Thus it is computationally burdensome and inconvenient to be applied in practice.

Since the infinitesimal operator has a closed-form in terms of drift and diffusion terms, the martingale based identification of diffusion process proposed here has the potential to do the separate inference in order to explore possible sources of the rejection. For simplicity, I only consider the univariate diffusion process and the extension to multivariate case is straightforward. Suppose $\{X_t\}$ follows a univariate diffusion model given by $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ where W_t is a 1-dimensional standard Brownian motion. By (2.18), the identification of the diffusion process is equivalent to a martingale property:

$$M_t^x = X_t - X_0 - \int_0^t b(X_s)ds \tag{4.1}$$

[14] Like Hong and Li(2005), my test procedure also enjoys the appealing property that a $\sqrt{n}$−consistent estimator is enough and the sampling variation in $\widehat{\theta}$ has no impact on the asymptotic distribution of $\widehat{M}_0(p)$. See the discussion in Section 4 for details.

and

$$M_t^{x^2} = X_t^2 - X_0^2 - 2\int_0^t b(X_s)X_s ds - \int_0^t \sigma^2(X_s)ds \tag{4.2}$$

are both martingales.

Observe that the first transformed process M_t^x only involves the drift term and the second $M_t^{x^2}$ has both the drift and diffusion terms as inputs. Intuitively, M_t^x characterizes the dynamics of the drift term solely and this characterization is robust to the dynamics of diffusion term. Note also that $\int_0^t \sigma^2(X_s)ds$ is the so-called "integrated volatility" or the quadratic variation $[X,X]_t$ of the process $\{X_t\}$ which has received extremely intensive attention in recent years(see Andersen, Bollerslev, Diebold, and Labys, 2003; Barndorff-Nielsen and Shephard 2004, 2006; Ait-Sahalia, Mykland and Zhang 2005;). Therefore $M_t^{x^2}$ contains the dynamics of diffusion term, i.e., the volatility of the process illustrated by $\int_0^t \sigma^2(X_s)ds$. Furthermore, $M_t^{x^2}$ also characterizes the interaction between drift and diffusion terms which is represented by $\int_0^t b(X_s)X_s ds$ because $b(X_s)X_s$ will raise the power of X_s at least to 2 and hence variance will also appear in this term.

Since the characterization for the dynamics of the drift term by the martingale property of M_t^x is robust to the dynamics of diffusion term, it is conceivable that we can check the specification of the drift term robust to diffusion misspecification if we further assume the drift term is identified by this characterization (4.1). Explicitly, suppose $\{X_t\}$ follows a univariate diffusion model given by $dX_t = b(X_t,\theta)dt + \sigma(X_t)dW_t$ with $\theta \in \Theta$ and Θ a finite dimensional parameter space, then the identification assumption is

Assumption 4.1: There exists a unique $\theta_0 \in \Theta$ such that $M_t^x(\theta) = X_t - X_0 - \int_0^t b(X_s;\theta)ds$ is a martingale.

Actually, under this assumption(this is equivalent to Assumption 2.1 in Park(2008)), Park(2008) proposes a so-called "conditional mean model of instantaneous change for a given stochastic process"[15] which is exactly the same as M_t^x here. The difference is that his model does not consider diffusion term at all and hence it can allow a more general setup, for example, jump diffusion process and stochastic volatility models. However, Park(2008) only proposes the instantaneous conditional mean model and claims that his model covers the diffusion process as a special case while he does not provide the corresponding conditions. Suppose the underlying model is a diffusion process and we are interested in testing the specification of the drift term. Then a test based on checking the maringale property of $M_t^x(\theta)$ is not omnibus. Since the identification not only involves $M_t^x(\theta)$ in (4.1) but also involves $M_t^{x^2}$ in (4.2), it could be the case that $M_t^x(\theta)$ is a martingale but $M_t^{x^2}(\theta, \sigma(\cdot))$ is not. In such a case, the test procedure only checking the maringale property of $M_t^x(\theta)$ cannot reject the null hypothesis although it should be rejected. This under-rejection may lead to misleading conclusion about the specification of diffusion models. In other words, Assumption 4.1 may be too restricted as an identification assumption and may not hold in many cases if a diffusion model is considered as the underlying process.

Observe that the martingale identification of drift here is based on rigourous mathematical derivation. Therefore, my infinitesimal operator based martingale characterizaion actually provides the mathematical conditions that the instantaneous conditional mean in Park's(2008) model is equal to the drift a diffusion

[15] Be careful about these terminologies. The instantaneous conditional mean for continuous time stochastic processes are different from the conditional mean for discrete time models. As discussed earlier, for instance, in a general diffusion process, the conditional mean of $X_{t+\Delta}$ given X_t is usually a function not of drift solely but of both drift and diffusion terms jointly. See Ait-Sahalia(1996a) and discussions below (2.12) in this paper.

process. If there is no information about whether the underlying process is a diffusion or not, Park's(2008) model is more general while if the diffusion model is regarded as the data generating process, the infinitesimal operator based martingale characterization should be considered. Moreover Park's(2008) identification of drift can be regarded as a special case of the infinitesimal operator based martingale characterization in the case of diffusion processes. The reson is that (4.1) which is also Park's(2008) identification assumption is derived using a special choice of function forms(see Theorem2 and discussion therein for details). If interesting function forms other than $f(x) = x_i$ and $x_i x_j$ in Theorem2 are suitably chosen, we may get other convenient and intutive characterizations of diffusion processes. Of course, as claimed by Park(2008), his model is a general conditional mean model of instantaneous change for continuous time stochastic processes. This makes his study more applicable in certain sense.

As a consequence, by assuming the drift term is identified by the martingale property of M_t^x, i.e., Assumption 4.1, a specification test for drift term can be constructed which is robust to diffusion term misspecification. The null hypothesis is the correct specification of drift term:

$$H_0 : P[b(X_t, \theta_0) = b^0(X_t)] = 1, \text{ for some } \theta_0 \in \Theta \text{ where } b^0(\cdot) \text{ is the true drift function}$$

which is equivalent to

$$H_0 : \text{For some } \theta_0 \in \Theta,\ M_t^x = X_t - X_0 - \int_0^t b(X_s;\theta)ds \text{ is a martingale} \tag{4.3}$$

Following the same reasoning as that for (3.2), I can test H_0 in (4.3) by checking the following $m.d.s.$property:

$$E\left[Y_{\tau\Delta}(\theta_0)|\mathcal{I}_{\tau-1}^Y\right] = 0, \text{ where } \mathcal{I}_{\tau-1}^Y = \sigma\{Y_{(\tau-1)\Delta}(\theta_0), Y_{(\tau-2)\Delta}(\theta_0), \cdots, Y_\Delta(\theta_0), Y_0(\theta_0)\}$$

where

$$Y_{\tau\Delta}(\theta_0) = X_{\tau\Delta} - X_{(\tau-1)\Delta} - \int_{(\tau-1)\Delta}^{\tau\Delta} b(X_s;\theta_0)ds \tag{4.4}$$

Let $Y_\tau(\theta_0) \equiv Y_{\tau\Delta}(\theta_0)$ for the simplification of notations. Obvisouly, $Y_\tau(\theta_0)$ cannot be observed. We first estimate the parameter θ_0 by the random sample $\{X_{\tau\Delta}\}_{\tau=1}^n$ to get a $\sqrt{n}$-consistent estimator and then the estimated processes $\widehat{Y}_\tau = Y_\tau(\widehat{\theta})$ is obtained. Since we are only interested in the specification of drift, it is better for us to use an estimation method which can estimate the parameters in the drift consistently while being robust to the diffusion misspecification. This essentially requries the estimation of a semi-parametric diffusion model with diffusion term unrestricted. Kristensen(2008a) and Ait-Sahalia(1996a) are examples of such methods. The test for checking (4.4) is a univariate special case of (3.9), i.e.,

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

where

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

$$\widehat{D}_0(p) = 2\sum_{m=1}^{n-2}\sum_{l=1}^{n-2} k^2(m/p)k^2(l/p)\int\int\left|\frac{1}{n-\max(m,l)}\sum_{\tau=\max(m,l)+1}^{n}\widehat{Y}_\tau^2\widehat{\psi}_{\tau-m}(v)\widehat{\psi}_{\tau-l}(u)\right|^2 dW(v)dW(u) \quad (4.5)$$

and $\widehat{\psi}_\tau(v) = e^{iv\widehat{Y}_\tau} - \widehat{\varphi}(v)$, and $\widehat{\varphi}(v) = n^{-1}\sum_{\tau=1}^{n} e^{iu\widehat{Y}_\tau}$ and all the terms are defined correspondingly for univariate case similar to multivariate case. Throughout, all unspecified integrals are taken on the support of $W(\cdot)$.

5 Asymptotic theory

5.1 Asymptotic distribution

Let

$$g^i(\tau,\theta) = -\int_{(\tau-1)\Delta}^{\tau\Delta} b_i(X_s;\theta)ds \tag{5.1}$$

and

$$g^{i,j}(\tau,\theta) = -\int_{(\tau-1)\Delta}^{\tau\Delta}\left[b_i(X_s;\theta)X_s^j + b_j(X_s;\theta)X_s^i + \sum_{k=1}^{d}[\sigma_{i,k}(X_s;\theta)\sigma_{j,k}(X_s;\theta)]\right] ds \tag{5.2}$$

Then we have

$$Z_\tau^i(\theta) = X_{\tau\Delta}^i - X_{(\tau-1)\Delta}^i + g^i(\tau,\theta) \tag{5.3}$$

and

$$Z_\tau^{i,j}(\theta) = X_{\tau\Delta}^i X_{\tau\Delta}^j - X_{(\tau-1)\Delta}^i X_{(\tau-1)\Delta}^j + g^{i,j}(\tau,\theta) \tag{5.4}$$

To derive the null asymptotic distribution the test statistic $\widehat{M}_0(p)$ in equation (3.9), the following regularity conditions are imposed.

Assumption A.1. $\{X_t\}$ is a strictly stationary time series such that $\mu = E[X_t]$ exists $a.s.$, and $E[\|Z_\tau\|^4] \leqq C$.

Assumption A.2. For each sufficiently large q, there exists a strictly stationary process $\{Z_{q,\tau}\}$ measurable with respect to the sigma field generated by $\{Z_{\tau-1}, Z_{\tau-2}, \cdots, Z_{\tau-q}\}$ such that as $q \to \infty$, $Z_{q,\tau}$ is independent of $\{Z_{\tau-q-1}, Z_{\tau-q-2}, \cdots\}$ for each τ, $E[Z_{q,\tau} \mid \mathcal{I}_{\tau-1}] = 0, a.s.$where $\mathcal{I}_{\tau-1}$ is the information set at time $(\tau-1)\Delta$ that may contain lagged random variables $\{X_{(\tau-m)\Delta}, m > 0\}$ from original process and lagged random variables $\{Z_{(\tau-m)\Delta}, m > 0\}$ from the transformed process, $E\|Z_\tau - Z_{q,\tau}\|^2 \leqq Cq^{-\kappa}$ for some constant $\kappa \geqq 1$, and $E\|Z_{q,\tau}\|^4 \leqq C$ for all large q.

Assumption A.3. With probability one, both $g^i(\tau,\cdot)$ and $g^{i,j}(\tau,\cdot)$ are continuously twice differentiable with respect to $\theta \in \Theta$ and $E\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}g^i(\tau,\theta)\right\|^4 \leqq C$, $E\sup_{\theta\in\Theta}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}g^i(\tau,\theta)\right\|^2 \leqq C$, $E\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}g^{i,j}(\tau,\theta)\right\|^4 \leqq$

C, and $E\sup_{\theta\in\Theta}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}g^{i,j}(\tau,\theta)\right\|^2 \leqq C$.

Assumption A.4. $\widehat{\theta}-\theta_0 = O_p(n^{-1/2})$, where $\theta_0 = p\lim(\widehat{\theta}) \in \Theta$.

Assumption A.5. $k:\mathbb{R}\rightarrow[-1,1]$ is symmetric and is continuous at $(0,0)$ and all but a finite number of points, with $k(0)=1$ and $|k(z)| \leqq C\,|z|^{-b}$ for large z and some $b>1$.

Assumption A.6. $W:\mathbb{R}^{d'}\rightarrow\mathbb{R}^+$ is nondecreasing and weighs sets symmetric about zero equallly, with $\int\|v\|^4\,dW(v)\leqq C$

Assumption A.7. Put $\psi_\tau(v) = e^{iv'Z_\tau}-\varphi(v)$ with $\varphi(v) = E\left[e^{iv'Z_\tau}\right]$ and $\sigma(a,a') = E\left(Z_\tau^a Z_\tau^{a'}\right)$ for $a,a' = i, ij$ and $i,j = 1,\cdots,d$(Note here ij does not denote the product between i and j but an index equivalent to (i,j). This notation applies to the whole paper). Then $\{\frac{\partial}{\partial\theta}g^i(\tau,\theta_0), Z_\tau\}$ and $\{\frac{\partial}{\partial\theta}g^{i,j}(\tau,\theta_0), Z_\tau\}$ are strictly stationary processes such that:

(a): $\sum_{m=1}^{\infty}\left\|Cov[\frac{\partial}{\partial\theta}g^a(\tau,\theta_0), \frac{\partial}{\partial\theta}g^a(\tau-m,\theta_0)]\right\| \leqq C$ for $a = i,(i,j)$ and $i,j=1,\cdots,d$;

(b): $\sum_{m=1}^{\infty}\sup_{(u,v)\in\mathbb{R}^{2d'}}|\sigma_m(u,v)| \leqq C$;

(c): $\sum_{m=1}^{\infty}\sup_{v\in\mathbb{R}^{d'}}\left\|Cov[\frac{\partial}{\partial\theta}g^a(\tau,\theta_0),\psi_{\tau-m}(v)]\right\| \leqq C$ for $a = i,(i,j)$ and $i,j=1,\cdots,d$;

(d): $\sum_{m,l=1}^{\infty}\sup_{(u,v)\in\mathbb{R}^{d'}}\left|E[(\left(Z_{\tau,a}Z_{\tau,a'}\right)-\sigma(a,a'))\psi_{\tau-m}(u)\psi_{\tau-l}(v)]\right| \leqq C$ for $a,a' = i,(i,j)$ and $i,j = 1,\cdots,d$;

(e)$\sum_{m,l,r=-\infty}^{\infty}\sup_{v\in\mathbb{R}^{d'}}\|\kappa_{m,l,r}(v)\|\leqq C$, where $\kappa_{m,l,r}(v)$ is the fourth order cumulant of the joint distribution of the process

$$\{\frac{\partial}{\partial\theta}g^a(\tau,\theta_0),\psi_{\tau-m}(v),\frac{\partial}{\partial\theta}g^a(\tau-l,\theta_0),\psi^*_{\tau-r}(v)\} \tag{5.5}$$

for $a = i, ij$ and $i,j = 1,\cdots,d$.

Assumptions A.1 and A.2 are regularity conditions on the data generating process (DGP). The strict stationarity on $\{X_t\}$ is imposed and the existence of the first order moment μ can be ensured by assuming $E\|X_t\|^2<\infty$. Assumption A.2 is required only under H_0. It assumes that the martingale difference sequence $(m.d.s.)$ $\{Z_\tau\}$ can be approximated by a $q-$dependent $m.d.s.$process $\{Z_{q,\tau}\}$ arbitrarily well when q is sufficiently large. Because $\{Z_\tau\}$ is a $m.d.s.$, Assumption A.2 essentially imposes restrictions on the serial dependence in higher order moments of X_τ. Besides, it implies ergodicity for $\{Z_\tau\}$. It holds trivially when $\{Z_\tau\}$ is a $q-$dependent process with an arbitrarily large but finite order q. In fact, this is general enough to cover many interesting processes, for example, a stochastic volatility model with short memory(see Hong and Lee(2005) for details).

Although Assumption A.3 appears in terms of restrictions on $g^i(\tau,\cdot)$ and $g^{i,j}(\tau,\cdot)$, it is actually imposing moment regularity conditions on the drift and diffusion terms $b(X_\tau;\theta_0)$ and $\sigma(X_\tau;\theta_0)$ which can be seen from (5.1) and (5.2). It covers most of the popular univariate and multivariate diffusion processes in both time-homogeneous and time-inhomogeneous cases, for example, Ait-Sahalia(1996a), Ahn and Gao(1999), Chan, Karolyi, Longstaff and Sanders(1992), Cox, Ingersoll and Ross(1985), and Vasicek(1977).

Assumption A.4 requires a $\sqrt{n}$-consistent estimator $\widehat{\theta}$, which may not be asymptotically most efficient. We do not need to know the asymptotic expansion of $\widehat{\theta}$, because the sampling variation in $\widehat{\theta}$ does not affect the limit distributions of $\widehat{M}_0(p)$. This delivers a convenient and generally applicable procedure in practice, because asymptotically most efficient estimators such as MLE or approximated MLE may be difficult to obtain in practice. One could choose a suboptimal, but convenient, estimator in implementing our procedure.

Assumption A.5 is a regularity condition on the kernel $k(\cdot)$. It contains all commonly used kernels in practice. The condition of $k(0) = 1$ ensures that the asymptotic bias of the smoothed kernel estimator $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ in (3.5) vanishes as $n \to \infty$. The tail condition on $k(\cdot)$ requires that $k(z)$ decays to zero sufficiently fast as $|z| \to \infty$. It implies $\int_0^\infty (1+z)k^2(z)dz < \infty$. For kernels with bounded support, such as the Bartlett and Parzen kernels, $b = \infty$. For the Daniell and quadratic-spectral kernels, $b = 1$ and 2, respectively. These two kernels have unbounded support, and thus all $(n-1)$ lags contained in the sample are used in constructing our test statistics. Assumption A.6 is a condition on the weighting function $W(\cdot)$ for the transform parameter v. It is satisfied by the CDF of any symmetric continuous distribution with a finite fourth moment. Finally, Assumption A.7 provides some covariance and fourth order cumulant conditions on $\{\frac{\partial}{\partial\theta}g^i(\tau,\theta_0), Z_\tau\}$ and $\{\frac{\partial}{\partial\theta}g^{i,j}(\tau,\theta_0), Z_\tau\}$, which restrict the degree of the serial dependence in $\{\frac{\partial}{\partial\theta}g^i(\tau,\theta_0), Z_\tau\}$ and $\{\frac{\partial}{\partial\theta}g^{i,j}(\tau,\theta_0), Z_\tau\}$. These conditions can be ensured by imposing more restrictive mixing and moment conditions on these two processes. However, to cover a sufficiently large class of DGPs, I choose not to do so.

I now state the asymptotic distribution of the test statistic $\widehat{M}_0(p)$ under H_0.

Theorem3: Suppose that Assumptions A.1-A.7 hold, and $p = cn^\lambda$ for $c \in (0,\infty)$ and $\lambda \in (0, \left(3 + \frac{1}{4b-2}\right)^{-1})$. Then under H_0,

$$\widehat{M}_0(p) \to^d N(0,1) \text{ as } n \to \infty$$

As an important feature of $\widehat{M}_0(p)$, the use of the estimated processes $\{\widehat{Z}_\tau\}$ in place of the true processes $\{Z_\tau\}$ has no impact on the limit distribution of $\widehat{M}_0(p)$. One can proceed as if the true parameter value θ_0 were known and equal to $\widehat{\theta}$. The reason, as pointed out by Hong and Lee(2005), is that the convergence rate of the parametric parameter estimator $\widehat{\theta}$ to θ 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 $\widehat{f}^{(0,1,0)}(\omega,0,v)$ and replacing θ_0 by $\widehat{\theta}$ has no impact asymptotically. This delivers a convenient procedure, because no specific estimation method for θ_0 is required[16]. Of course, parameter estimation uncertainty in $\widehat{\theta}$ may have impact on the small sample distribution of $\widehat{M}_0(p)$. In small samples, one can use a bootstrap procedure to obtain more accurate levels of the tests.

5.2 Asymptotic power

My tests are derived without assuming an alternative model to H_0. To gain insight into the nature of the alternatives that my tests are able to detect, I now examine the asymptotic behavior of $\widehat{M}_0(p)$ under H_A. For

[16] Because of the nice properties just discussed, $\widehat{M}_0(p)$ can be used to test the *m.d.s.*hypothesis for the process with conditional heteroscedasticity of unknown form. As discussed in Hong and Lee(2005), Lobato (2002) and Park and Whang (2003) proposed some nonparametric tests of the *m.d.s.* for observed raw data using the conditioning indicator function. They also allowed for conditional heteroscedasticity, and Park and Whang (2003) allowed for nonstationary conditioning variables. However, these tests only check a fixed lag order. Moreover, their limit distributions depend on the DGP and cannot be tabulated;resampling methods have to be used to obtain critical values on a case-by-case basis. That is why I choose to extend Hong's(1999) generalized spectral approach instead of using these methods.

this purpose, I impose a condition on the serial dependence in $\{Z_\tau\}$:

Assumption A.8. $\sum_{m=1}^{\infty} \sup_{v\in\mathbb{R}^{d'}} \left|\sigma_m^{(1,0)}(0,v)\right| \leqq C.$

Theorem4: Suppose Assumptions A.1 and A.3-A.8 hold, and $p = cn^{\lambda}$ for $c \in (0,\infty)$ and $\lambda \in (0, 1/2)$. Then under H_A and as $n \to \infty$,

$$\begin{aligned}(p^{1/2}/n)\widehat{M}_0(p) &\to^{p} \left[2D\int_0^{\infty} k^4(z)dz\right]^{-1/2} \int \left\|f^{(0,1,0)}(\omega,0,v) - f_0^{(0,1,0)}(\omega,0,v)\right\|^2 d\omega dW(v) \\ &= \left[2D\int_0^{\infty} k^4(z)dz\right]^{-1/2} \sum_{m=1}^{\infty} \int \left\|\sigma_m^{(1,0)}(0,v)\right\|^2 dW(v)\end{aligned}$$

where

$$D = 2\sum_{a=1}^{d'}\sum_{a'=1}^{d'} E\left|Z_{a\tau}Z_{a'\tau}\right| \int\int\int_{-\pi}^{\pi} |f(\omega,u,v)|^2 d\omega dW(u)dW(v) \tag{5.6}$$

The constant D takes into account the impact of the serial dependence in conditioning variables $\{e^{iv'Z_{\tau-m}}\}$, which generally exists even under H_0, due to the presence of the serial dependence in the conditional variance and higher order moments of $\{Z_\tau\}$. This differs from the $i.i.d.$case, where we can show that D depends only on the marginal distribution of $\{Z_\tau\}$.

Suppose the autoregression function $E[Z_\tau \mid Z_{\tau-m}] \neq 0$ 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 $E[Z_\tau \mid Z_{\tau-m}] \neq 0$ at some lag $m > 0$. The main advantage of $\widehat{M}_0(p)$ is that it can eventually detect all possible misspecifications of $m.d.s$ property that render $E[Z_\tau \mid Z_{\tau-m}] \neq 0$ at some lag $m > 0$. This avoids the blindness of searching for different alternatives when one has no prior information.

5.3 Data-Driven Lag order

A practical issue in implementing our tests is the choice of the lag order p. As an advantage, the smoothing generalized spectral approach can provide a data-driven method to choose p, which, to some extent, lets data themselves speak for a proper p. Before discussing any specific method, I first justify the use of a data-driven lag order, $\widehat{p}$, say. Here, we impose a Lipschitz continuity condition on $k(\cdot)$. This condition rules out the truncated kernel $k(z) = 1(|z| \leqq 1)$, but it still contains most commonly used nonuniform kernels.

Assumption A.9. $|k(z_1) - k(z_2)| \leqq C\,|z_1 - z_2|$ for any (z_1, z_2) in $\mathbb{R}^2$ and some constant $C < \infty$.

Theorem5: Suppose that Assumptions A.1-A.7 and A.9 hold, and $\widehat{p}$ is a data-driven bandwidth such that $\widehat{p}/p = 1 + O_p(p^{-(\frac{3}{2}\beta-1)})$ for some $\beta > \frac{2b-1/2}{2b-1}$, where b is as in Assumption A.5, and p is a nonstochastic bandwidth with $p = cn^{\lambda}$ for $c \in (0, \infty)$ and $\lambda \in (0, \left(3 + \frac{1}{4b-2}\right)^{-1})$. Then under H_0,

$$\widehat{M}_0(\widehat{p}) - \widehat{M}_0(p) \rightarrow^p 0 \text{ and } \widehat{M}_0(\widehat{p}) \rightarrow^d N(0,1)$$

Hence, the use of $\widehat{p}$ has no impact on the limit distribution of $\widehat{M}_0(\widehat{p})$ as long as $\widehat{p}$ converges to p sufficiently fast and my test procedure enjoys an additional "nuisance parameter-free" property. Theorem6 allows for a wide range of admissible rates for $\widehat{p}$. One possible choice is the nonparametric plug-in method similar to Hong (1999, Theorem 2.2) which minimizes an asymptotic integrated mean square error(IMSE) criterion for the estimator $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ in (3.5). Consider some "pilot" generalized spectral derivative estimators based on a preliminary bandwidth $\overline{p}$:

$$\overline{f}^{(0,1,0)}(\omega, 0, v) = \frac{1}{2\pi} \sum_{m=1-n}^{n-1} (1 - |m|/n)^{1/2} \overline{k}(m/\overline{p}) \widehat{\sigma}_m^{(1,0)}(0, v) e^{-im\omega}$$

$$\overline{f}^{(q,1,0)}(\omega, 0, v) = \frac{1}{2\pi} \sum_{m=1-n}^{n-1} (1 - |m|/n)^{1/2} \overline{k}(m/\overline{p}) \widehat{\sigma}_m^{(1,0)}(0, v) \, |m|^q \, e^{-im\omega} \tag{5.7}$$

where the kernel $\overline{k}(\cdot)$ needs not be the same as the kernel $k(\cdot)$ used in (3.5). Note that $\overline{f}^{(0,1,0)}(\omega, 0, v)$ is an estimator for $f^{(0,1,0)}(\omega, 0, v)$ and $\overline{f}^{(q,1,0)}(\omega, 0, v)$ is an estimator for the generalized spectral derivative

$$f^{(q,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sigma_m^{(1,0)}(0, v) \, |m|^q \, e^{-im\omega} \tag{5.8}$$

For the kernel $k(\cdot)$, suppose there exists some $q \in (0, \infty)$ such that

$$0 < k^{(q)} = \lim_{z \to 0} \frac{1 - k(z)}{|z|^q} \tag{5.9}$$

Then I define the plug-in bandwidth as

$$\widehat{p}_0 = \widehat{c}_0 n^{\frac{1}{2q+1}} \tag{5.10}$$

where the turning parameter estimator

$$\begin{aligned}
\widehat{c}_0 &= \left\{ \frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty} k^2(z) dz} \frac{\int \int_{-\pi}^{\pi} \left\| \overline{f}^{(q,1,0)}(\omega, 0, v) \right\|^2 d\omega dW(v)}{\int_{-\pi}^{\pi} \left\| \int \overline{f}^{(0,1,0)}(\omega, v, -v) dW(v) \right\|^2 d\omega} \right\}^{\frac{1}{2q+1}} \\
&= \left\{ \frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty} k^2(z) dz} \frac{\sum_{m=1-n}^{n-1} (n - |m|) \overline{k}^2(m/\overline{p}) \, |m|^{2q} \int \left\| \widehat{\sigma}_m^{(1,0)}(0, v) \right\|^2 dW(v)}{\sum_{m=1-n}^{n-1} (n - |m|) \overline{k}^2(m/\overline{p}) \widehat{R}(m) \int \| \widehat{\sigma}_m(v, -v) \| \, dW(v)} \right\}^{\frac{1}{2q+1}}
\end{aligned} \tag{5.11}$$

and $\widehat{R}(m) = (n - |m|)^{-1} \sum_{\tau=|m|+1}^{n} \widehat{Z}'_{\tau} \widehat{Z}_{\tau-|m|}$.

The data-driven $\widehat{p}_0$ in (5.10) involves the choice of a preliminary bandwidth $\widehat{p}$, which can be fixed or grow with the sample size n. If it is fixed, $\widehat{p}_0$ still generally grows at rate $n^{\frac{1}{2q+1}}$ under H_A, but $\widehat{c}_0$ does not converge to the optimal tuning constant c_0 (say) that minimizes the IMSE of $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ in (3.5). This is a parametric plug-in method. Alternatively, following Hong (1999), we can show that when $\overline{p}$ grows with n properly, the data-driven bandwidth $\widehat{p}_0$ in (5.10) will minimize an asymptotic IMSE of $\widehat{f}^{(0,1,0)}(\omega, 0, v)$. The choice of $\overline{p}$ is somewhat arbitrary, but we expect that it is of secondary importance[17].

6 Finite Sample Performance

6.1 Empirical Size of the Test

I now study the finite sample performance of the test procedure. The simulation design is the same as those in Hong and Li (2005) and Pritsker (1998). To examine the size of the test, I simulate data from Vasicek's (1977) model:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t \tag{6.1}$$

where α is the long run mean and κ is the speed of mean reversion. The smaller κ is, the stronger the serial dependence in $\{X_t\}$, and consequently, the slower the convergence to the long run mean. I am particularly interested in the possible impact of dependent persistence in $\{X_t\}$ on the size of the test. Given that the finite sample performance of the test may depend on both the marginal density and dependent persistence of $\{X_t\}$, I follow Hong and Li (2005) and Pritsker(1998) to change κ and σ^2 in the same proportion so that the marginal density is unchanged;namely,

$$p(x, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left[-\frac{(x-\alpha)^2}{2\sigma_s^2}\right]$$

where the stationary variance $\sigma_s^2 = \sigma^2/(2\kappa) = 0.01226$. In this way, we can focus on the impact of dependent persistence. We consider both low and high levels of dependent persistence and adopt the same parameter values as Hong and Li (2005) and Pritsker (1998): $\left(\kappa, \alpha, \sigma^2\right) = (0.85837, 0.089102, 0.002185)$ and $(0.214592, 0.089102, 0.000546)$ for the low and high persistent dependence cases respectively.

For each parameterization, we simulate 1,000 data sets of a random sample $\{X_{\tau\Delta}\}_{\tau=1}^{n}$ at the monthly frequency for $n = 250, 500$,and 1000 respectively. The simulation is carried out based on the transition density of $\{X_t\}$ which is normal with mean $X_t \exp(-\kappa\Delta) + \alpha\left[1 - \exp(-\kappa\Delta)\right]$ and the variance $\sigma^2\left[1 - \exp(-2\kappa\Delta)\right]/(2\kappa)$. The initial value X_0 is generated from the marginal density $p(x, \boldsymbol{\theta})$. These sample sizes correspond to about $20 - 100$ years of monthly data. For each data set, we estimate the model parameters $\theta = (\kappa, \alpha, \sigma^2)'$ via the MLE method and compute the statistics based on the estimated Vasicek model. I consider the empiri-

[17] Hong and Lee(2005) pointed out that the data-driven $\widehat{p}$ based on the IMSE criterion generally will not maximize the power of $\widehat{M}_0(p)$. They suggested that a more sensible alternative may be to develop a data-driven $\widehat{p}$ using a power criterion, or a criterion that trades off level distortion and power loss. But they did not pursue that method and content with the IMSE criterion because it will necessitate higher order asymptotic analysis. Also their simulation experience suggests that the power of our tests seems to be relatively flat in the neighbourhood of the optimal lag order that maximizes the power, and $\widehat{p}_0$ in (5.10) performs reasonably well in finite samples.

cal rejection rates using the asymptotic critical values (1.28 and 1.65) at the 10% and 5% significance levels respectively.

The Bartlett kernel is used both in computing the data-dependent optimal bandwidth $\widehat{p}_0$ by the plug-in method for some preliminary bandwidth $\overline{p}$ and in computing the test statistic $\widehat{M}_0(\widehat{p}_0)$. We choose the standard multivariate normal CDF for $W(\cdot)$. Our simualtion experience indicates that choices of kernel function $k(\cdot)$ and weight function $W(\cdot)$ have no substantial impact on the size performance of tests. Table I reports the empirical sizes of $\widehat{M}_0(\widehat{p}_0)$ at the 10% and 5% levels under a correct Vasicek model. Both of the cases with low and high persistence of dependence are considered. It can be observed that there is over-rejection for both cases at both 10% and 5% levels, but the performances are improving as n increases. Since the over-rejection is still serious especially at the 5% level when $n = 1000$, I increase the sample size to $n = 1500$(only for size performance) to check the empirical sizes of the test. Obviously, when the sample size is large enough, the test has good size performances with a bit over-rejection at the 5% level which is not very serious. Furthermore, the tests display more over-rejections under strong mean reversion than under weak mean reversion, which is similar to Chen and Hong's(2008a) conditional characteristic function based test. Another observation worth pointing out is that the rejection rates do not vary much for different choices of preliminary bandwidths. This can be seen as a robust property of the optimal bandwidth based on plug-in methods.

6.2 Empirical Power of the Test

To investigate the power of the test, we simulate data from four popular diffusion models other than Vasicek model and test the null hypothesis that data are generated from a Vasicek model.

- DGP1 (CIR Model):

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t \tag{6.2}$$

where $(\kappa, \alpha, \sigma^2) = (0.89218,\ 0.090495,\ 0.032742)$.

- DGP2 (Ahn and Gao's (1999) Inverse-Feller Model):

$$dX_t = X_t[\kappa - \left(\sigma^2 - \kappa\alpha\right) X_t]dt + \sigma X_t^{3/2} dW_t \tag{6.3}$$

where $(\kappa, \alpha, \sigma^2) = (3.4387,\ 0.0828,\ 1.420864)$.

- DGP3 (CKLS (Chan, Karolyi, Longstaff and Sanders, 1992) Model):

$$dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{\rho} dW_t \tag{6.4}$$

where $(\kappa, \alpha, \sigma^2, \rho) = (0.0972,\ 0.0808,\ 0.52186,\ 1.46)$.

- DGP4 (Ait-Sahalia's (1996a) Nonlinear Drift Model):

$$dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)dt + \sigma X_t^{\rho} dW_t \tag{6.5}$$

where $((\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma^2, \rho)$=(0.00107,-0.0517, 0.877, -4.604, 0.64754, 1.50).

Follwing Hong and Li(2005), the parameter values for the CIR model are taken from Pritsker (1998), and the parameter values for Ahn and Gao's inverse-Feller model from Ahn and Gao (1999)[18]. For DGPs 3 and 4, the parameter values are taken from Ait-Sahalia's (1999) estimates of real interest rate data. For each of these four alternatives, we generate 1000 data sets of the random sample for $\{X_\tau\}_{\tau=\Delta}^{n\Delta}$ where $n = 250, 500$ and 1000 respectively at the monthly sample frequency. Simulated data for CIR and Ahn and Gao's model are based on closed-form model transition densities. For the CKLS and Ait-Sahalia's nonlinear drift models, whose transition densities have no closed-form expressions, we simulate data by the Euler-Milstein scheme. Each simulated sample path is generated using 40 intervals per month with 39 discarded out of every 40 observations, obtaining discrete observations at the monthly frequency.

The Vasicek model (6.1), which is implied by the null hypothesis H_0, is estimated by MLE for each data set. Table II reports the rejection rates of $\widehat{M}_0(\widehat{p}_0)$ at the 10% and 5% levels using empirical critical values, which are obtained under H_0. Under DGP1, Vasicek model is correctly specified for the drift function but is misspecified for the diffusion function because it fails to capture the "level effect". The test has good power in this case, with rejection rates around over 96% at the 5% level when $n = 1000$. Under DGP2, Vasicek model is misspecified for both the instantaneous mean–drift and the instantaneous variance–diffusion because it ignores the nonlinear drift and diffusion. As expected, the test $\widehat{M}_0(\widehat{p}_0)$ has excellent power when Vasicek model (6.1) is used to fit the data generated from DGP2. The power increases substantially with the sample size n and approaches unity when $n = 1000$.

Similar to DGP1, DGP3 is only mis-specified for the diffusion term, with the only difference that the coefficient of elasticity for volatility is equal to 1.46 now. The rejection rates are low when the sample size n is 250, but increases very quickly to over 55% when $n = 500$ and to over 95% when $n = 1000$. Under DGP4, Vasicek model (6.1) is misspecified for both the drift and diffusion terms because it ignores the nonlinearity in both terms. The rejection rates are already over 80% at the 5% level when $n = 500$. To sum up, the proposed test has good omnibus power in detecting various model mis-specifications. The power performances are reasonable even when the sample size n is only 500.

7 Conclusion

I develop an omnibus specification test for diffusion models based on the infinitesimal operator instead of the already extensively used transition density. The infinitesimal operator-based identification of the diffusion process is equivalent to a "martingale hypothesis" for the new processes transformed from the original diffusion process by the celebrated "martingale problems". My test procedure is to check the "martingale hypothesis" by extending Hong's(1999) generalized spectral approach to a multivariate generalized spectral derivative based test which has many good properties. The infinitesimal operator of the diffusion process enjoys the nice property of being a closed-form expression of drift and diffusion terms. This makes my test procedure capable of checking both univariate and multivariate diffusion models and particularly powerful and convenient for

[18] Chen and Hong(2008a) found some typos in the parameter values of Ahn and Gao's(1999) inverse-feller model by private correspondence and corrected therm. Here I choose the parameter values used by them.

the multivariate case while in contrast checking the multivariate diffusion models is very difficult by transition density-based methods because transition density does not have a closed-form in general.

Moreover, different transformed martingale processes based on infinitesimal operator and "martingale problems" contain different separate information about the drift and diffusion terms or their interactions. This motivates us to discuss several feasible test procedures which are to do separate inference to explore the sources when rejection of a parametric form happens. Finally, simulation studies show that the proposed tests have reasonable size performances and excellent power performances in finite sample. Possible future researches about diffusion processes using the infinitesimal operator-based martingale characterization discussed in the paper are being pursued.

A drawback of the infinitesimal operator based identification is that it only holds for the pure diffusion process and will fail when the sample path of the process exhibits discontinuities, the so-called “jumps”, my test procedure actually rules out jumps a priori. This is somewhat unsatisfactory in practice especially for high frequency data for which jumps are now believed to be an essential component of asset price dynamics both emprically and theoretically(Ait-Sahalia 2002a; Barndorff-Nielsen and Shephard 2004, 2006; Lee and Mykland 2008; Ait-Sahalia and Jacod 2008; Andersen et al. 2002; Johannes 2004; and Pan 2002). In this case, we can frist identify and then discard the jump points from the sample path using methods in Lee and Mykland(2008), Andersen, Bollerslev and Dobrev(2007), Fan and Fan(2008), and Fan and Wang(2007). This makes the test procedure proposed here more applicable and robust.

Mathematical Appendix

Throughout the Appendix, let $g_i(\tau,\theta)$, $g_{ij}(\tau,\theta)$, $Z^i_\tau(\theta)$, and $Z^{ij}_\tau(\theta)$ be defined as in (5.1)-(5.4). I let $M_0(p)$ be defined in the same way as $\widehat{M}_0(p)$ in (3.5) with the unobservable sample $\{Z_\tau = Z_{\tau\Delta}(\theta_0)\}_{\tau=1}^n$, where $\theta_0 = p\lim\widehat{\theta}$, replacing the estimated processes samples $\{\widehat{Z}_\tau = Z_{\tau\Delta}(\widehat{\theta})\}_{\tau=1}^n$. Also, $C \in (1,\infty)$ denotes a generic bounded constant.

Proof of Theorem 1. See ChV.19-20 of Rogers and Williams(2000), or Theorem 21.7 of Kallengber(2002), or Proposition 2.4 of ChVII in Revuz and Yor(2005). ■

Proof of Theorem 2. See Proposition 4.6 and Remark 4.12 of Karatzas and Shreve(1991, Ch5.4) ■

Proof of Theorem 3. It suffices to show Theorems A1-A3 below. Theorem A1 implies that replacing $\{Z_\tau\}_{\tau=1}^n$ by $\{\widehat{Z}_\tau\}_{\tau=1}^n$ has no impact on the limit distribution of $\widehat{M}_0(p)$; Theorem A2 says that the use of truncated process $\{Z_{q,\tau}\}_{\tau=1}^n$ rather than the original $\{Z_\tau\}_{\tau=1}^n$ does not affect the limit distribution of $\widehat{M}_0(p)$ for q sufficiently large. The assumption that $Z_{q,\tau}$ is independent of $\{Z_{\tau-m}\}_{m=q+1}^\infty$ when q is large greatly simplifies the derivation of asymptotic normality of $\widehat{M}_0(p)$. ■

TheoremA1. Under the conditions of Theorem 3, $\widehat{M}_0(p) - M_0(p) \to^p 0$.

TheoremA2. Let $M_{0q}(p)$ be defined as $M_0(p)$ with $\{Z_{q,\tau}\}_{\tau=1}^n$ replacing $\{Z_\tau\}_{\tau=1}^n$,where $\{Z_{q,\tau}\}$ is as in Assumption A.2. Then under the conditions of Theorem3 and $q = p^{1+\frac{1}{4b-2}}\left(\ln^2 n\right)^{\frac{1}{2b-1}}$, $M_{0q}(p) - M_0(p) \to^p 0$.

TheoremA3. Under the conditions of Theorem3 and $q = p^{1+\frac{1}{4b-2}}\left(\ln^2 n\right)^{\frac{1}{2b-1}}$, $M_{0q}(p) \to^d N(0,1)$.

Proof of Theorem A1. Note that $Z_\tau(\theta)$ has components $Z_\tau^i(\theta) = X_{\tau\Delta}^i - X_{(\tau-1)\Delta}^i + g_i(\tau,\theta)$ and $Z_\tau^{ij}(\theta) = X_{\tau\Delta}^i X_{\tau\Delta}^j - X_{(\tau-1)\Delta}^i X_{(\tau-1)\Delta}^j + g_{ij}(\tau,\theta)$ and similarly $\widehat{Z}_\tau$ has components $\widehat{Z}_\tau^i = X_{\tau\Delta}^i - X_{(\tau-1)\Delta}^i + g_i(\tau,\widehat{\theta})$ and $\widehat{Z}_\tau^{ij} = X_{\tau\Delta}^i X_{\tau\Delta}^j - X_{(\tau-1)\Delta}^i X_{(\tau-1)\Delta}^j + g_{ij}(\tau,\widehat{\theta})$ respectively. By the mean value theorem, we have $\widehat{Z}_\tau^i = Z_\tau^i - g_i^{'}(\tau,\overline{\theta})'(\widehat{\theta} - \theta_0)$ for some $\overline{\theta}$ between $\widehat{\theta}$ and θ_0, where $g_i^{'}(\tau,\theta) = \frac{\partial}{\partial\theta} g_i(\tau,\theta)$. By the Cauchy-Schwartz inequality and Assumptions A.3-A.4,

$$\sum_{\tau=1}^{n}\left(\widehat{Z}_\tau^i - Z_\tau^i\right)^2 \leqq n\left\|\widehat{\theta} - \theta_0\right\|^2 n^{-1}\sum_{\tau=1}^{n}\sup_{\theta\in\Theta_0}\left\|g_i'(\tau,\theta)\right\|^2 = O_p(1), \text{ for any } i = 1,\cdots,d \tag{A.1}$$

where Θ_0 is a neighbourhood of θ_0. By similar reasoning, we have

$$\sum_{\tau=1}^{n}\left(\widehat{Z}_\tau^{ij} - Z_\tau^{ij}\right)^2 \leqq n\left\|\widehat{\theta} - \theta_0\right\|^2 n^{-1}\sum_{\tau=1}^{n}\sup_{\theta\in\Theta_0}\left\|g_i'(\tau,\theta)\right\|^2 = O_p(1), \text{ for any } i,j = 1,\cdots,d \tag{A.2}$$

(A.1) and (A.2) together imply that

$$\begin{aligned}\sum_{\tau=1}^{n}\left\|\widehat{Z}_\tau - Z_\tau\right\|^2 &= \sum_{\tau=1}^{n}\left[\sum_{i=1}^{d}\left(\widehat{Z}_\tau^i - Z_\tau^i\right)^2 + \sum_{i=1}^{d}\sum_{j=1}^{d}\left(\widehat{Z}_\tau^{ij} - Z_\tau^{ij}\right)^2\right] \\ &= \sum_{i=1}^{d}\left[\sum_{\tau=1}^{n}\left(\widehat{Z}_\tau^i - Z_\tau^i\right)^2\right] + \sum_{i=1}^{d}\sum_{j=1}^{d}\left[\sum_{\tau=1}^{n}\left(\widehat{Z}_\tau^{ij} - Z_\tau^{ij}\right)^2\right] \\ &= O_p(1) \end{aligned}\tag{A.3}$$

Now put $n_m = n - |m|$, and let $\widetilde{\sigma}_m^{(1,0)}(0,v)$ be defined in the same way as $\widehat{\sigma}_m^{(1,0)}(0,v)$ in (3.5), with $\{Z_\tau\}_{\tau=1}^n$ replacing $\{\widehat{Z}_\tau\}_{\tau=1}^n$. To show $\widehat{M}_0(p) - M_0(p) \to^p 0$, it is sufficient to prove

$$\widehat{D}_0^{-\frac{1}{2}}(p)\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left[\left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2 - \left\|\widetilde{\sigma}_m^{(1,0)}(0,v)\right\|^2\right]dW(v) \to^p 0 \tag{A.4}$$

$\widehat{C}_0(p) - \widetilde{C}_0(p) = O_p(n^{-1/2})$, and $\widehat{D}_0(p) - \widetilde{D}_0(p) = o_p(1)$, where $\widetilde{C}_0(p)$ and $\widetilde{D}_0(p)$ are defined in the same way as $\widehat{C}_0(p)$ and $\widehat{D}_0(p)$ in (3.9) with $\{Z_\tau\}_{\tau=1}^n$ replacing $\{\widehat{Z}_\tau\}_{\tau=1}^n$. To save space, I focus on the proof of (A.4); the proofs for $\widehat{C}_0(p) - \widetilde{C}_0(p) = O_p(n^{-1/2})$, and $\widehat{D}_0(p) - \widetilde{D}_0(p) = o_p(1)$ are routine. Note that it is necessary to acchieve the convergence rate $O_p(n^{-1/2})$ for $\widehat{C}_0(p) - \widetilde{C}_0(p)$ to make sure that repalcing $\widehat{C}_0(p)$ with $\widetilde{C}_0(p)$ has asymptotically negligible impact given $p/n \to 0$.

Since

$$
\begin{aligned}
&\left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2 - \left\|\widetilde{\sigma}_m^{(1,0)}(0,v)\right\|^2 \\
&= \sum_{i=1}^{d}\left[\left|\widehat{\sigma}_{m,i}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2\right] + \sum_{i=1}^{d}\sum_{j=1}^{d}\left[\left|\widehat{\sigma}_{m,ij}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,ij}^{(1,0)}(0,v)\right|^2\right] \qquad \text{(A.5)}
\end{aligned}
$$

where $\widehat{\sigma}_{m,i}^{(1,0)}(0,v)$ and $\widehat{\sigma}_{m,ij}^{(1,0)}(0,v)$ for $i,j=1,\cdots,d$ are the components of $\widehat{\sigma}_m^{(1,0)}(0,v)$ and correspondingly $\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)$ and $\widetilde{\sigma}_{m,ij}^{(1,0)}(0,v)$ for $i,j=1,\cdots,d$ are the components of $\widetilde{\sigma}_m^{(1,0)}(0,v)$. By (A.5), it is sufficient for (A.4) to show that

$$
\widehat{D}_0^{-\frac{1}{2}}(p)\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left[\left|\widehat{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 - \left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2\right]dW(v)\to^p 0,\ \text{for } i=1,\cdots,d \qquad \text{(A.6)}
$$

and

$$
\widehat{D}_0^{-\frac{1}{2}}(p)\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left[\left|\widehat{\sigma}_{m,ij}^{(1,0)}(0,v)\right|^2 - \left|\widetilde{\sigma}_{m,ij}^{(1,0)}(0,v)\right|^2\right]dW(v)\to^p 0,\ \text{for } i,j=1,\cdots,d \qquad \text{(A.7)}
$$

We will only show (A.6) here and the proof of (A.7) is similar.

To show (A.6), I first decompose

$$
\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left[\left|\widehat{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 - \left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2\right]dW(v) = \widehat{A}_1 + 2\,\mathrm{Re}(\widehat{A}_2) \qquad \text{(A.8)}
$$

where

$$
\widehat{A}_1 = \int\sum_{m=1}^{n-1}k^2(m/p)n_m\left|\widehat{\sigma}_{m,i}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 dW(v)
$$

$$
\widehat{A}_2 = \int\sum_{m=1}^{n-1}k^2(m/p)n_m\left[\widehat{\sigma}_{m,i}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right]\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)^* dW(v)
$$

where $\mathrm{Re}(\widehat{A}_2)$ denote the real part of $\widehat{A}_2$ and $\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)^*$ denote the complex congugate of $\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)$. Then (A.6) follows from the following Propositions A1 and A2 and $p\to\infty$ as $n\to\infty$. ■

PropositionA1. Under the conditions of Theorem3, $\widehat{A}_1 = O_p(1)$.

PropositionA2. Under the conditions of Theorem3, $p^{-1/2}\widehat{A}_2 = o_p(1)$.

Proof of Proposition A1. Put $\widehat{\delta}_\tau(v) = e^{iv'\widehat{Z}_\tau} - e^{iv'Z_\tau}$ and $\psi_\tau(v) = e^{iv'Z_\tau} - \varphi(v)$, where $\varphi(v) = Ee^{iv'Z_\tau}$. Then straightforward algebra yields that for $m>0$,

$$
\begin{aligned}
&\widehat{\sigma}_{m,i}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,i}^{(1,0)}(0,v) \\
&\qquad = in_m^{-1} \sum_{\tau=m+1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i})\widehat{\delta}_{\tau-m}(v) - i \left[n_m^{-1} \sum_{\tau=m+1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i}) \right] \left[n_m^{-1} \sum_{\tau=m+1}^{n} \widehat{\delta}_{\tau-m}(v) \right] \\
&\qquad\quad + in_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i}\widehat{\delta}_{\tau-m}(v) - i \left[n_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i} \right] \left[n_m^{-1} \sum_{\tau=m+1}^{n} \widehat{\delta}_{\tau-m}(v) \right] \\
&\qquad\quad + in_m^{-1} \sum_{\tau=m+1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i})\psi_{\tau-m}(v) - i \left[n_m^{-1} \sum_{\tau=m+1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i}) \right] \left[n_m^{-1} \sum_{\tau=m+1}^{n} \psi_{\tau-m}(v) \right] \\
&\qquad = i \left[\widehat{B}_{1m}(v) - \widehat{B}_{2m}(v) + \widehat{B}_{3m}(v) - \widehat{B}_{4m}(v) + \widehat{B}_{5m}(v) - \widehat{B}_{6m}(v) \right], \text{ say} \qquad \text{(A.9)}
\end{aligned}
$$

Then it follows that $\widehat{A}_1 \leqq 8 \sum_{a'=1}^{6} \sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{a'm}(v) \right|^2 dW(v)$. Proposition A1 follows from Lemmas A1-A6 below and $p/n \to 0$. ■

LemmaA1. $\sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{1m}(v) \right|^2 dW(v) = O_p(p/n)$.

LemmaA2. $\sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{2m}(v) \right|^2 dW(v) = O_p(p/n)$.

LemmaA3. $\sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{3m}(v) \right|^2 dW(v) = O_p(p/n)$.

LemmaA4. $\sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{4m}(v) \right|^2 dW(v) = O_p(p/n)$.

LemmaA5. $\sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{5m}(v) \right|^2 dW(v) = O_p(1)$.

LemmaA6. $\sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{6m}(v) \right|^2 dW(v) = O_p(p/n)$.

Now let $a_n(m) = n_m^{-1} k^2(m/p)$. In the following, I will show these lemmas above.

Proof of Lemma A1. By the Cauchy-Schwartz inequality and the inequality that $\left| e^{iz_1} - e^{iz_2} \right| \leqq |z_1 - z_2|$ for any real-valued variables z_1 and z_2, I have

$$
\left| \widehat{B}_{1m}(v) \right|^2 \leqq \left[n_m^{-1} \sum_{\tau=1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i})^2 \right] \left[n_m^{-1} \sum_{\tau=1}^{n} \left| \widehat{\delta}_\tau(v) \right|^2 \right] \leqq \|v\|^2 \left[n_m^{-1} \sum_{\tau=1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i})^2 \right]^2
$$

It follows from (A.1) and Assumptions A.5-A.6 that

$$
\int \sum_{m=1}^{n-1} k^2(m/p) n_m^{-1} \left| \widehat{B}_{1m}(v) \right|^2 dW \leqq \left[\sum_{m=1}^{n-1} a_n(m) \right] \left[\sum_{\tau=1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i})^2 \right]^2 \int \|v\|^2 \, dW(v) = O_p(p/n) \qquad \text{(A.10)}
$$

where I made use of the fact that

$$\sum_{m=1}^{n-1} a_n(m) = \sum_{m=1}^{n-1} n_m^{-1} k^2(m/p) = O(p/n) \tag{A.11}$$

given $p = cn^\lambda$ for $\lambda \in (0, 1/2)$, as shown in Hong(1999, A.15, Page 1213). ■

Proof of Lemma A2. By the inequality that $\left|e^{iz_1} - e^{iz_2}\right| \leqq |z_1 - z_2|$ for any real-valued variables z_1 and z_2, I have

$$\left|\widehat{B}_{2m}(v)\right|^2 \leqq \left[n_m^{-1} \sum_{\tau=1}^{n} \left|\widehat{Z}_{\tau,i} - Z_{\tau,i}\right|\right]^2 \left[n_m^{-1} \sum_{\tau=1}^{n} \left|v\widehat{Z}_{\tau,i} - vZ_{\tau,i}\right|\right]^2 \leqq \|v\|^2 \left[n_m^{-1} \sum_{\tau=1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i})^2\right]^2$$

By the same reasoning as that of Lemma A1, the desired result follows. ■

Proof of Lemma A3. Using the inequality that $\left|e^{iz} - 1 - iz\right| \leqq |z|^2$ for any real-valued variables z, I have

$$\left|e^{iv'\widehat{Z}_{\tau-m}} - e^{iv'Z_{\tau-m}} - iv\left[\widehat{Z}_{\tau-m,i} - Z_{\tau-m,i}\right] e^{iv'Z_{\tau-m}}\right| \leqq \|v\|^2 \left[\widehat{Z}_{\tau-m,i} - Z_{\tau-m,i}\right]^2 \tag{A.12}$$

A seond order Taylor series expansion yields

$$\widehat{Z}_{\tau-m,i} = Z_{\tau-m,i} - \left(g_i'(\tau - m, \theta_0)\right)' (\widehat{\theta} - \theta_0) - \frac{1}{2}(\widehat{\theta} - \theta_0)' g_i''(\tau - m, \overline{\theta})(\widehat{\theta} - \theta_0) \tag{A.13}$$

for some $\overline{\theta}$ between $\widehat{\theta}$ and θ_0, where $g_i''(\tau, \theta) \equiv \frac{\partial^2}{\partial\theta\partial\theta'} g(\tau, \theta)$. Put $\xi_\tau(v) = g_i'(\tau, \theta_0) e^{iv'Z_\tau}$. Then (A.12) and (A.3) imply that

$$\left|e^{iv'\widehat{Z}_{\tau-m}} - e^{iv'Z_{\tau-m}} - iv\xi_{\tau-m}(v)(\widehat{\theta} - \theta_0)\right| \leqq \|v\|^2 \left[\widehat{Z}_{\tau-m,i} - Z_{\tau-m,i}\right]^2 + \|v\| \left\|\widehat{\theta} - \theta_0\right\|^2 \sup_{\theta \in \Theta_0} \left\|g_i''(\tau - m, \theta)\right\|$$

where Θ_0 is a neighbourhood of θ_0.

Henceforth, by (A.9), I obtain

$$\begin{aligned} n_m \left|\widehat{B}_{3m}(v)\right| \leqq{}& \|v\| \left\|\widehat{\theta} - \theta_0\right\| \left\|\sum_{\tau=m+1}^{n} Z_{\tau,i}\xi_{\tau-m}(v)\right\| + \|v\|^2 \sum_{\tau=m+1}^{n} |Z_{\tau,i}| \left[\widehat{Z}_{\tau-m,i} - Z_{\tau-m,i}\right]^2 \\ &+ \|v\| \left\|\widehat{\theta} - \theta_0\right\|^2 \sum_{\tau=m+1}^{n} |Z_{\tau,i}| \sup_{\theta \in \Theta_0} \left\|g_i''(\tau - m, \theta)\right\| \end{aligned}$$

Then it follows from Assumptions A1-A7 and (A.11) that

$$
\begin{aligned}
& \sum_{m=1}^{n-1} \int k^2(m/p) n_m \left| \widehat{B}_{3m}(v) \right|^2 dW(v) \\
\leqq \; & 4 \left\| \sqrt{n}(\widehat{\theta} - \theta_0) \right\|^2 \sum_{m=1}^{n-1} k^2(m/p) \int \left\| n_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i} \xi_{\tau-m}(v) \right\|^2 \|v\|^2 dW(v) \\
& + 4 \left\| \sqrt{n}(\widehat{\theta} - \theta_0) \right\|^4 \left(n^{-1} \sum_{\tau=1}^{n} Z_{\tau,i}^2 \right) \left\{ n^{-1} \sum_{\tau=1}^{n} \left[\sup_{\theta \in \Theta_0} \left\| g_i'(\tau, \theta) \right\| \right]^4 \right\} \left[\sum_{m=1}^{n-1} a_n(m) \right] \int \|v\|^4 dW(v) \\
& + 4 \left\| \sqrt{n}(\widehat{\theta} - \theta_0) \right\|^4 \left(n^{-1} \sum_{\tau=1}^{n} Z_{\tau,i}^2 \right) \left\{ n^{-1} \sum_{\tau=1}^{n} \left[\sup_{\theta \in \Theta_0} \left\| g_i''(\tau, \theta) \right\| \right]^2 \right\} \left[\sum_{m=1}^{n-1} a_n(m) \right] \int \|v\|^2 dW(v) \\
= \; & O_p(p/n) \qquad \text{(A.14)}
\end{aligned}
$$

by the fact that $E \left\| \sum_{\tau=m+1}^{n} Z_{\tau,i} \xi_{\tau-m}(v) \right\|^2 \leqq C n_m$ given $E\left(Z_{\tau,i} \mid I_{\tau-1}\right) = 0$ *a.s.* under H_0 and Assumptions A.1 and A.3. ■

Proof of Lemma A4. By the Cauchy-Schwartz inequality,

$$
\left| \widehat{B}_{4m}(v) \right|^2 \leqq \left(n_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i} \right)^2 n_m^{-1} \sum_{\tau=m+1}^{n} \left| \widehat{\delta}_\tau(v) \right|
$$

Then by this inequality, Cauchy-Schwartz again, and $\left| \widehat{\delta}_\tau(v) \right| \leqq \left| v' \left(\widehat{Z}_\tau - Z_\tau \right) \right|$,

$$
\begin{aligned}
\sum_{m=1}^{n-1} \int k^2(m/p) n_m \left| \widehat{B}_{4m}(v) \right|^2 dW(v) & \leqq \sum_{m=1}^{n-1} k^2(m/p) \left(n_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i} \right)^2 \left[\sum_{\tau=1}^{n} \left\| \widehat{Z}_\tau - Z_\tau \right\|^2 \right] \int \|v\|^2 dW(v) \\
& = O_p(p/n)
\end{aligned}
$$

given (A.3) and (A.11), and $E\left(\sum_{\tau=m+1}^{n} Z_{\tau,i}\right)^2 = \sigma^2 n_m$ by H_0, the *m.d.s.* hypothesis of $\{Z_\tau\}$. ■

Proof of Lemma A5. By the second order Taylor series expansion in (A.13),

$$
-\widehat{B}_{5m}(v) = (\widehat{\theta} - \theta_0)' n_m^{-1} \sum_{\tau=m+1}^{n} g_i'(\tau, \theta_0) \psi_{\tau-m}(v) + \frac{1}{2} (\widehat{\theta} - \theta_0)' \left[n_m^{-1} \sum_{\tau=m+1}^{n} g_i''(\tau, \overline{\theta}) \psi_{\tau-m}(v) \right] (\widehat{\theta} - \theta_0)
$$

for some $\overline{\theta}$ between $\widehat{\theta}$ and θ_0. Then I have

$$
\begin{aligned}
& \sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{5m}(v) \right|^2 dW(v) \\
\leqq \;& 2 \left\| \sqrt{n}(\widehat{\theta} - \theta_0) \right\|^2 \sum_{m=1}^{n-1} k^2(m/p) \int \left\| n_m^{-1} \sum_{\tau=m+1}^{n} g_i'(\tau, \theta_0) \psi_{\tau-m}(v) \right\|^2 dW(v) \\
& +2 \left\| \sqrt{n}(\widehat{\theta} - \theta_0) \right\|^4 \left[n^{-1} \sum_{\tau=1}^{n} \sup_{\theta \in \Theta_0} \left\| g_i''(\tau, \theta) \right\| \right]^2 \left[\sum_{m=1}^{n-1} a_n(m) \right] \int dW(v) \\
= \;& O_p(1) + O_p(p/T) \qquad \text{(A.15)}
\end{aligned}
$$

where the last term is $O_p(p/T)$ given (A.11) and the first term is $O_p(1)$, as is shown below:

Put $\eta_m(v) = E\left(g_i'(\tau, \theta_0)\psi_{\tau-m}(v)\right) = Cov[g_i'(\tau, \theta_0), \psi_{\tau-m}(v)]$. Then

$$
\sup_{v \in \mathbb{R}^{d'}} \sum_{m=1}^{\infty} \|\eta_m(v)\| \leqq C
$$

by Assumption A.7. Then expressing the moments in terms of cumulants by the well-known formulas(see Hannan, 1970, (5.1), Page 23 for real-valued processes and also Stratonovich(1963), chapter 1 and Leonov and Shiryaev (1959) for more details), I can obtain

$$
\begin{aligned}
& n_m E \left\| n_m^{-1} \sum_{\tau=m+1}^{n} g_i'(\tau, \theta_0) \psi_{\tau-m}(v) - \eta_m(v) \right\|^2 \\
\leqq \;& \sum_{r=-n_m}^{n_m} \left\| Cov[g_i'(\tau, \theta_0), g_i'(-r, \theta_0)'] \right\| \cdot |\sigma_r(v, -v)| + \sum_{r=-n_m}^{n_m} \left\| \eta_{m+|r|}(-v) \right\| \cdot \left\| \eta_{m-|r|}(v) \right\| + \sum_{r=-n_m}^{n_m} \left\| \kappa_{m,|r|,m+|r|}(v) \right\| \\
\leqq \;& C \qquad \text{(A.16)}
\end{aligned}
$$

given Assumption A.7, where $\kappa_{m,l,r}(v)$ is as in Assumption A.7. As a result, from (A.11) and (A.16), $|k(\cdot)| \leqq 1$, and $p/n \to 0$, I get

$$
\begin{aligned}
& \sum_{m=1}^{n-1} k^2(m/p) E \int \left\| n_m^{-1} \sum_{\tau=m+1}^{n} g_i'(\tau, \theta_0) \psi_{\tau-m}(v) \right\|^2 dW(v) \\
\leqq \;& C \sum_{m=1}^{n-1} \int \|\eta_m(v)\|^2 dW(v) + C \sum_{m=1}^{n-1} a_n(m) = O(1) + O(p/n) = O(1)
\end{aligned}
$$

Therefore the first term in (A.10) is $O_p(1)$. ■

Proof of Lemma A6. The proof is analogous to that of Lemma A4. ■

Proof of Proposition A2. Given the decomposition in (A.9), I have

$$\left|\left[\widehat{\sigma}_{m,i}^{(1,0)}(0,v)-\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right]\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)^*\right| \leqq \sum_{a'=1}^{6}\left|\widehat{B}_{a'm}(v)\right|\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right| \tag{A.17}$$

where $\widehat{B}_{a'm}(v)$ is defined in (A.9). By the Cauchy-Schwartz inequality,

$$\begin{aligned}
&\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widehat{B}_{a'm}(v)\right|\cdot\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|dW(v)\\
\leqq\;&\left[\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widehat{B}_{a'm}(v)\right|^2dW(v)\right]^{1/2}\left[\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2dW(v)\right]^{1/2}\\
=\;&O_p(p^{1/2}/n^{1/2})O_p(p^{1/2})=o_p(p^{1/2}),\ a'=1,2,3,4,6,
\end{aligned}$$

given Lemmas A1-A4 and A6, and $p/n\to 0$, where

$$p^{-1}\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2dW(v)=O_p(1)$$

by Markov's inequality, the $m.d.s.$property of $\{Z_\tau\}$ under H_0, and (A.9).

Then consider the case $a^{'}=5$. By Assumptions A1-A7,

$$\begin{aligned}
&\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widehat{B}_{5m}(v)\right|\cdot\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|dW(v)\\
\leqq\;&\left\|\widehat{\theta}-\theta_0\right\|\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left\|n_m^{-1}\sum_{\tau=m+1}^{n}g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right\|\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|dW(v)\\
&+n\left\|\widehat{\theta}-\theta_0\right\|^2\left[n^{-1}\sum_{m=1}^{n-1}\sup_{\theta\in\Theta_0}\left\|g_i''(\tau,\theta)\right\|\right]\sum_{m=1}^{n-1}k^2(m/p)\int\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|dW(v)\\
=\;&O_p(1+p/n^{1/2})+O_p(p/n^{1/2})=o_p(p^{1/2})
\end{aligned}\tag{A.18}$$

given $p\to\infty$ and $p/n\to 0$, where I have used the fact that

$$n_mE\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2\leq C$$

by the $m.d.s.$ property of $\{Z_\tau\}$ under H_0 and the fact that the first term in (A.18) is $O_p(1+p/n^{1/2})$, as shown below:

By (A.16) and Cauchy-Schwartz inequality, I have

$$E\left[\left\|n_m^{-1}\sum_{\tau=m+1}^{n} g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right\|\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|\right]$$

$$\leqq \left[E\left\|n_m^{-1}\sum_{\tau=m+1}^{n} g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right\|^2\right]^{1/2}\left[E\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2\right]^{1/2} \leqq C\left[\|\eta_m(v)\|+Cn_m^{-1/2}\right]n_m^{-1/2}$$

and consequently

$$n^{-\frac{1}{2}}\sum_{m=1}^{n-1}k^2(m/p)n_m E\int\left\|n_m^{-1}\sum_{\tau=m+1}^{n} g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right\|\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|dW(v)$$

$$\leqq C\sum_{m=1}^{n-1}\int\|\eta_m(v)\|\,dW(v)+Cn^{-\frac{1}{2}}\sum_{m=1}^{n-1}k^2(m/p)=O(1+p/n^{1/2})$$

given $|k(\cdot)|\leqq 1$ and Assumption A.7. ■

Proof of Theorem A2. The proof is similar to Theorem A1. By the same reasoning as that of (A.4)-(A.7), we will consider only the case $i=1\cdots,d$. Let $\widehat{A}_{1,q}$ and $\widehat{A}_{2,q}$ be defined in the same way as $\widehat{A}_1$ and $\widehat{A}_2$ in (A.8), with $\{Z_{q,\tau}\}_{\tau=1}^n$ replacing $\{\widehat{Z}_\tau\}_{\tau=1}^n$. It is enough to show that $p^{-\frac{1}{2}}\widehat{A}_{1,q}\to^p 0$ and $p^{-\frac{1}{2}}\widehat{A}_{2,q}\to^p 0$.

Let $\delta_{q,\tau}=e^{iv'Z_\tau}-e^{iv'Z_{q,\tau}}$ and $\psi_{q,\tau}(v)=e^{iv'Z_{q,\tau}}-\varphi_q(v)$, where $\varphi_q(v)=E[e^{iv'Z_{q,\tau}}]$. Let $\widetilde{\sigma}_{q,m}^{(1,0)}(0,v)$ be defined as $\widetilde{\sigma}_m^{(1,0)}(0,v)$ with $\{Z_{q,\tau}\}_{\tau=1}^n$ replacing $\{Z_\tau\}_{\tau=1}^n$. Then similar to (A.9), I have

$$\begin{aligned}
&\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)-\widetilde{\sigma}_{q,m,i}^{(1,0)}(0,v)\\
&\quad= in_m^{-1}\sum_{\tau=m+1}^{n}(Z_{\tau,i}-Z_{q,\tau,i})\delta_{q,\tau-m}(v)-i\left[n_m^{-1}\sum_{\tau=m+1}^{n}(Z_{\tau,i}-Z_{q,\tau,i})\right]\left[n_m^{-1}\sum_{\tau=m+1}^{n}\delta_{q,\tau-m}(v)\right]\\
&\qquad+in_m^{-1}\sum_{\tau=m+1}^{n}Z_{q,\tau,i}\delta_{q,\tau-m}(v)-i\left[n_m^{-1}\sum_{\tau=m+1}^{n}Z_{q,\tau,i}\right]\left[n_m^{-1}\sum_{\tau=m+1}^{n}\delta_{q,\tau-m}(v)\right]\\
&\qquad+in_m^{-1}\sum_{\tau=m+1}^{n}(Z_{\tau,i}-Z_{q,\tau,i})\psi_{q,\tau-m}(v)-i\left[n_m^{-1}\sum_{\tau=m+1}^{n}(Z_{\tau,i}-Z_{q,\tau,i})\right]\left[n_m^{-1}\sum_{\tau=m+1}^{n}\psi_{q,\tau-m}(v)\right]\\
&\quad= i\left[\widehat{B}_{1mq}(v)-\widehat{B}_{2mq}(v)+\widehat{B}_{3mq}(v)-\widehat{B}_{4mq}(v)+\widehat{B}_{5mq}(v)-\widehat{B}_{6mq}(v)\right],\text{ say}
\end{aligned}$$

Following the same reasoning as that of Theorem A1 and noting that $E[Z_\tau\mid I_{\tau-1}]=0$ *a.s.* and $E[Z_{q,\tau}\mid I_{\tau-1}]=0$ *a.s.*, we have

$$p^{-\frac{1}{2}}\widehat{A}_{1,q}\leqq 8p^{-\frac{1}{2}}\sum_{a'=1}^{6}\sum_{\tau=1}^{n-1}k^2(m/p)n_m\int\left|\widehat{B}_{a'mq}(v)\right|^2dW(v)=O_p(p^{\frac{1}{2}}/q^{\kappa})=o_p(1)$$

given Assumption A.2, $q/p\to\infty$, and $\kappa\geqq 1$. Further, by Cauchy-Schwartz inequality,

$$p^{-\frac{1}{2}}\widehat{A}_{2,q}=2p^{-\frac{1}{2}}\sum_{a'=1}^{6}\sum_{\tau=1}^{n-1}k^2(m/p)n_m \operatorname{Re}\int \widehat{B}_{a'mq}(v)\widetilde{\sigma}_{q,m,i}^{(1,0)}(0,v)^* dW(v)=O_p(p^{\frac{1}{2}}/q^{\kappa})=o_p(1)$$

This completes the proof of Theorem A.2. ■

Proof of Theorem A3. I shall show Proposition A.3 and A.4 below. ■

Proposition A.3: Let $\widetilde{\sigma}_{q,m}^{(1,0)}(0,v)$ be defined as $\widetilde{\sigma}_{m}^{(1,0)}(0,v)$, and let $\widetilde{C}_{0q}(p)$ be defined as $\widetilde{C}_0(p)$, with $\{Z_{q,\tau}\}_{\tau=1}^n$ replacing $\{Z_\tau\}_{\tau=1}^n$. Then under the conditions of Theorem1,

$$p^{-1/2}\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left\|\widetilde{\sigma}_{q,m}^{(1,0)}(0,v)\right\|^2 dW(v)=p^{-1/2}\widetilde{C}_{0q}(p)+p^{-1/2}\widetilde{V}_q+o_p(1) \tag{A.19}$$

where

$$\widetilde{V}_q=\sum_{a=i \text{ and } (i,j) \text{ with } i,j=1,\cdots d}\widetilde{V}_{q,a} \text{ and } \widetilde{C}_{0q}(p)=\sum_{a=i \text{ and } (i,j) \text{ with } i,j=1,\cdots d}\widetilde{C}_{0q,a}(p)$$

and

$$\begin{aligned}
\widetilde{V}_{q,a} &= \sum_{\tau=2q+2}^{n} Z_{q,\tau,a}\sum_{m=1}^{q}a_n(m)\int \psi_{q,\tau-m}(v)\left[\sum_{s=1}^{\tau-2q-1}Z_{q,s,a}\psi_{q,s-m}^*(v)\right]dW(v)\\
\widetilde{C}_{0q,a}(p) &= \sum_{m=1}^{n-1}k^2(m/p)\frac{1}{n-m}\sum_{\tau=m+1}^{n-1}Z_{q,\tau,a}^2\int\left|\psi_{\tau-m}(v)\right|^2 dW(v)
\end{aligned}$$

Proposition A.4: Let $\widetilde{D}_{0q}(p)$ be defined as $\widetilde{D}_0(p)$ with $\{Z_{q,\tau}\}_{\tau=1}^n$ replacing $\{Z_\tau\}_{\tau=1}^n$. Then

$$\left[\widetilde{D}_{0q}(p)\right]^{-1/2}\widetilde{V}_q\to^d N(0,1)$$

Proof of Proposition A.3: Recall that $\widetilde{\sigma}_{q,m,a}^{(1,0)}(0,v)=n_m^{-1}\sum_{\tau=m+1}^{n}Z_{q,\tau,a}\psi_{q,\tau-m}(v)$, where $\psi_{q,\tau}(v)\equiv e^{iv'Z_{q,\tau}}-\varphi_q(v)$ and $\varphi_q(v)=E\left(e^{iv'Z_{q,\tau}}\right)$. Then

$$\begin{aligned}
\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left\|\widetilde{\sigma}_{q,m}^{(1,0)}(0,v)\right\|^2 dW(v) &= \sum_{m=1}^{n-1}k^2(m/p)n_m\int\sum_{a=i \text{ and } (i,j) \text{ with } i,j=1,\cdots d}\left|\widetilde{\sigma}_{q,m,a}^{(1,0)}(0,v)\right|^2 dW(v)\\
&= \sum_{a=i \text{ and } (i,j) \text{ with } i,j=1,\cdots d}\left[\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widetilde{\sigma}_{q,m,a}^{(1,0)}(0,v)\right|^2 dW(v)\right]
\end{aligned}$$

Henceforth, to prove (A.19), it is sufficient to show that

$$p^{-1/2}\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widetilde{\sigma}_{q,m,a}^{(1,0)}(0,v)\right|^2 dW(v)=p^{-1/2}\widetilde{C}_{0q,a}(p)+p^{-1/2}\widetilde{V}_{q,a}+o_p(1) \tag{A.20}$$

To show (A.20), I first decompose

$$
\begin{aligned}
&\sum_{m=1}^{n-1} k^2(m/p)n_m \int \left|\widetilde{\sigma}_{q,m,a}^{(1,0)}(0,v)\right|^2 dW(v) \\
=\ & \sum_{m=1}^{n-1} a_n(m) \int \left|\sum_{\tau=1}^{n} Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right|^2 dW(v) + \sum_{m=1}^{n-1} a_n(m) \int \left|\sum_{\tau=1}^{m} Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right|^2 dW(v) \\
& -2\,\mathrm{Re}\sum_{m=1}^{n-1} a_n(m) \int \left[\sum_{\tau=1}^{n} Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right]\left[\sum_{\tau=1}^{m} Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right]^* dW(v \\
\equiv\ & \widetilde{Q}_q + \widetilde{R}_{1q} - 2\,\mathrm{Re}\left(\widetilde{R}_{2q}\right) \qquad \text{(A.21)}
\end{aligned}
$$

Next write

$$
\begin{aligned}
\widetilde{Q}_q &= \sum_{m=1}^{n-1} a_n(m) \int \sum_{\tau=1}^{n} Z_{q,\tau,a}^2 \left|\psi_{q,\tau-m}(v)\right|^2 dW(v) + 2\,\mathrm{Re}\sum_{m=1}^{n-1} a_n(m) \int \sum_{\tau=2}^{n}\sum_{s=1}^{\tau-1} Z_{q,\tau,a}Z_{q,s,a}\psi_{q,\tau-m}(v)\psi_{q,s-m}^*(v)dW(v) \\
&\equiv \widetilde{C}_q(p) + 2\,\mathrm{Re}\left(\widetilde{U}_q\right) \qquad \text{(A.22)}
\end{aligned}
$$

and further decompose

$$
\begin{aligned}
\widetilde{U}_q &= \sum_{\tau=2q+2}^{n} Z_{q,\tau,a} \int \sum_{m=1}^{n-2} a_n(m)\psi_{q,\tau-m}(v) \sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi_{q,s-m}^*(v)dW(v) \\
&\quad + \sum_{\tau=2}^{n} Z_{q,\tau,a} \int \sum_{m=1}^{n-2} a_n(m)\psi_{q,\tau-m}(v) \sum_{s=\max(1,\tau-2q)}^{\tau-1} Z_{q,s,a}\psi_{q,s-m}^*(v)dW(v) \\
&\equiv \widetilde{U}_{1q} + \widetilde{R}_{3q} \qquad \text{(A.23)}
\end{aligned}
$$

where in the first term $\widetilde{U}_{1q}$, we have $\tau - s > 2q$ so that $\{Z_{q,\tau,a}, \psi_{q,\tau-m}(v)\}_{m=1}^q$ is independent of $\{Z_{q,s,a}, \psi_{q,s-m}(v)\}_{m=1}^q$ for q sufficiently large. In the second term $\widetilde{R}_{3q}$, we have $0 < \tau - s \leq 2q$. Finally, write

$$
\begin{aligned}
\widetilde{U}_{1q} &= \sum_{\tau=2q+2}^{n} Z_{q,\tau,a} \sum_{m=1}^{q} a_n(m) \int \psi_{q,\tau-m}(v) \sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi_{q,s-m}^*(v)dW(v) \\
&\quad + \sum_{\tau=2q+2}^{n} Z_{q,\tau,a} \sum_{m=q+1}^{n-1} a_n(m) \int \psi_{q,\tau-m}(v) \sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi_{q,s-m}^*(v)dW(v) \\
&\equiv \widetilde{V}_{q,a} + \widetilde{R}_{4q} \qquad \text{(A.24)}
\end{aligned}
$$

where the first term $\widetilde{V}_{q,a}$ is contributed by the lad orders m from 1 to q; and the second term $\widetilde{R}_{4q}$ is from lag orders $m > q$. It follows from (A.21) to (A.24) that

$$
\sum_{m=1}^{n-1} k^2(m/p)n_m \int \left|\widetilde{\sigma}_{q,m,a}^{(1,0)}(0,v)\right|^2 dW(v) = \widetilde{C}_q(p) + 2\,\mathrm{Re}\left(\widetilde{V}_{q,a}\right) + \widetilde{R}_{1q} - 2\,\mathrm{Re}\left(\widetilde{R}_{2q} - \widetilde{R}_{3q} - \widetilde{R}_{4q}\right)
$$

It sufficies to show Lemmas A.7-A.11 below, which imply $p^{-1/2}\left[\widetilde{C}_q(p)-\widetilde{C}_{0q,a}(p)\right]=o_p(1)$ and $p^{-1/2}\widetilde{R}_{a'q}=o_p(1)$ for $a'=1,2,3,4$ given $q=p^{1+\frac{1}{4b-2}}\left(\ln^2 n\right)^{\frac{1}{2b-1}}$ and $p=cn^{\lambda}$ for $0<\lambda<\left(3+\frac{1}{4b-2}\right)^{-1}$. ■

Lemma A.7: Let $\widetilde{C}_q(p)$ be defined as in (A.22). Then $\widetilde{C}_q(p)-\widetilde{C}_{0q,a}(p)=O_p(p^2/n)$.

Lemma A.8: Let $\widetilde{R}_{1q}$ be defined as in (A.21). Then $\widetilde{R}_{1q}=O_p(p^2/n)$.

Lemma A.9: Let $\widetilde{R}_{2q}$ be defined as in (A.21). Then $\widetilde{R}_{2q}=O_p(p^{\frac{3}{2}}/n^{\frac{1}{2}})$.

Lemma A.10: Let $\widetilde{R}_{3q}$ be defined as in (A.23). Then $\widetilde{R}_{3q}=O_p(q^{\frac{1}{2}}p/n^{\frac{1}{2}})$.

Lemma A.11: Let $\widetilde{R}_{4q}$ be defined as in (A.24). Then $\widetilde{R}_{4q}=O_p(p^{2b}\ln(n)/q^{2b-1})$.

Proof of Lemma A.7: By Markov's inequality and $E\left|\widetilde{C}_q(p)-\widetilde{C}_{0q,a}(p)\right|\leq Cp^2/n$ given $\sum_{m=1}^{n-1}(m/p)\,a_n(m)=O(p/n)$. ■

Proof of Lemma A.8: By the *m.d.s.* property of $\{Z_{q,\tau},\mathcal{F}_{\tau-1}\}$ where $\mathcal{F}_{\tau-1}$ is the sigma-field generated by $\{Z_{\tau-m}\}_{m=1}^{\infty}$, we can obtain $E\int\left|\sum_{\tau=1}^{m}Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right|^2dW(v)=\sum_{\tau=1}^{m}\int E\left[Z_{q,\tau,a}^2\left|\psi_{q,\tau-m}(v)\right|^2\right]dW(v)\leq Cm$. The result then follows from Markov's inequality and $\sum_{m=1}^{n-1}(m/p)\,a_n(m)=O(p/n)$ given Assumption A.6. ■

Proof of Lemma A.9: The proof is similar to that of Lemma A.8, with the fact that

$$E\left|\int\left[\sum_{\tau=1}^{n}Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right]\left[\sum_{\tau=1}^{m}Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right]^{*}dW(v)\right|\leq C\,(mn)^{1/2}$$

given Assumption A.6. ■

Proof of Lemma A.10: By the *m.d.s.* property of $\{Z_{q,\tau},\mathcal{F}_{\tau-1}\}$, Minkowski's inequality and (A.11), we have

$$\begin{aligned}
E\left|\widetilde{R}_{3q}\right|^2 &= \sum_{\tau=2}^{n}E\left|\sum_{m=1}^{n-1}a_n(m)\int Z_{q,\tau,a}\psi_{q,\tau-m}(v)\sum_{s=\max(1,\tau-2q)}^{\tau-1}Z_{q,s,a}\psi_{q,s-m}^{*}(v)dW(v)\right|^2\\
&\leq \sum_{\tau=2}^{n}\left[\sum_{m=1}^{n-1}a_n(m)\int\left(E\left|Z_{q,\tau,a}\psi_{q,\tau-m}(v)\sum_{s=\max(1,\tau-2q)}^{\tau-1}Z_{q,s,a}\psi_{q,s-m}^{*}(v)\right|^2\right)^{\frac{1}{2}}dW(v)\right]^2\\
&\leq 2Cnq\left[\sum_{m=1}^{n-1}a_n(m)\right]^2=O(qp^2/n)
\end{aligned}$$

This finishes the proof of Lemma A.10. ■

Proof of Lemma A.11: By the $m.d.s.$ property of $\{Z_{q,\tau}, \mathcal{F}_{\tau-1}\}$ and Minkowski's inequality,

$$
\begin{aligned}
E\left|\widetilde{R}_{4q}\right|^2 &= \sum_{\tau=2q+2}^{n} E\left|\sum_{m=q+1}^{n-1} a_n(m)\int Z_{q,\tau,a}\psi_{q,\tau-m}(v)\sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi^*_{q,s-m}(v)dW(v)\right|^2 \\
&\leq \sum_{\tau=2q+2}^{n}\left[\sum_{m=q+1}^{n-1} a_n(m)\int\left(E\left|Z_{q,\tau,a}\psi_{q,\tau-m}(v)\sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi^*_{q,s-m}(v)\right|^2\right)^{\frac{1}{2}} dW(v)\right]^2 \\
&\leq Cn^2\left[\sum_{m=q+1}^{n-1} a_n(m)\right]^2 \leq Cn^2\left[\sum_{m=q+1}^{n-1}(m/p)^{-2b}\, n_m^{-1}\right]^2 = O(p^{4b}\ln^2(n)/q^{4b-2})
\end{aligned}
$$

given that fact that $k(z) \leq C\,|z|^{-b}$ as $z \to \infty$ from Assumption A.6. ■

Proof of Proposition A.4: From (A.19), $\widetilde{V}_q = \sum_{a=i \text{ and } (i,j) \text{ with } i,j=1,\cdots d} \widetilde{V}_{q,a}$. We rewrite $\widetilde{V}_q = \sum_{\tau=2q+2}^{n} V_q(\tau)$, where

$$
V_q(\tau) = \sum_{a=i \text{ and } (i,j) \text{ with } i,j=1,\cdots d} V_{q,a},\; V_{q,a} = Z_{q,\tau,a}\sum_{m=1}^{q} a_n(m)\int \psi_{q,\tau-m}(v) H_{m,.\tau-2q-1,a}(v)dW(v)
$$

and

$$
H_{m,.\tau-2q-1,a}(v) = \sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi^*_{q,s-m}(v)
$$

Then I will apply the martingale central limit theorem(Brown, 1971), which states that $var\left(2\operatorname{Re}\widetilde{V}_q\right)^{-\frac{1}{2}} 2\operatorname{Re}\widetilde{V}_q \to^d N(0,1)$ if

$$
var\left(2\operatorname{Re}\widetilde{V}_q\right)^{-1}\sum_{\tau=1}^{n}\left[2\operatorname{Re}V_q(\tau)\right]^2 1\left[\left|2\operatorname{Re}V_q(\tau)\right| > \eta\cdot var\left(2\operatorname{Re}\widetilde{V}_q\right)^{\frac{1}{2}}\right] \to 0,\ \forall\eta > 0 \qquad \text{(A.25)}
$$

$$
var\left(2\operatorname{Re}\widetilde{V}_q\right)^{-1}\sum_{\tau=1}^{n} E\left[2\operatorname{Re}V_q^2(\tau)|\mathcal{F}_{\tau-1}\right] \to^p 1 \qquad \text{(A.26)}
$$

First, let's compute $var\left(2\operatorname{Re}\widetilde{V}_q\right)^{-1}$. By the $m.d.s.$ property of $\{Z_{q,\tau}, \mathcal{F}_{\tau-1}\}$ under H_0 and independence between $Z_{q,\tau}$ and $\{Z_{\tau-m-1}\}_{m=q}^{\infty}$ for q sufficiently large, we have

$$
\begin{aligned}
& E\left(\widetilde{V}_q^2\right) \\
= & \sum_{\tau=2q+2}^{n} \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} E\left[Z_{q,\tau,a}Z_{q,\tau,a'} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\right. \\
& \left. \int\int \psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)H_{m,.\tau-2q-1,a}(v)H_{m,.\tau-2q-1,a'}(u)dW(v)dW(u)\right] \\
= & \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int \sum_{\tau=2q+2}^{n} \sum_{s=1}^{\tau-2q-1} \\
& E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] E\left[Z_{q,s,a}Z_{q,s,a'}\psi^*_{q,s-m}(v)\psi^*_{q,s-l}(u)\right] dW(v)dW(u) \\
= & \frac{1}{2} \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} k^2(m/p)k^2(l/p) \int\int \left|E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m}(v)\psi_{q,-l}(u)\right]\right|^2 dW(v)dW(u)\left[1+o(1)\right]
\end{aligned}
$$

Similarly, we can obtain

$$
\begin{aligned}
& E\left(\widetilde{V}_q^*\right)^2 \\
= & \frac{1}{2} \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} k^2(m/p)k^2(l/p) \int\int \left|E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m}(v)\psi_{q,-l}(u)\right]\right|^2 dW(v)dW(u)\left[1+o(1)\right]
\end{aligned}
$$

and

$$
\begin{aligned}
& E\left|\widetilde{V}_q\right|^2 \\
= & \frac{1}{2} \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} k^2(m/p)k^2(l/p) \int\int \left|E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m}(v)\psi_{q,-l}(u)\right]\right|^2 dW(v)dW(u)\left[1+o(1)\right]
\end{aligned}
$$

Because $W(\cdot)$ wights sets symmetric about zero equally, we have $E\left|\widetilde{V}_q\right|^2 = E\left(\widetilde{V}_q^2\right) = E\left(\widetilde{V}_q^*\right)^2$. Hence

$$
\begin{aligned}
& var\left(2\operatorname{Re}\widetilde{V}_q\right) \\
& = 2E\left|\widetilde{V}_q\right|^2 + E\left(\widetilde{V}_q^2\right) + E\left(\widetilde{V}_q^*\right)^2 \\
& = 2\sum\sum_{a,a'=i,\ (i,j),\ i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} k^2(m/p)k^2(l/p) \int\int \left|E\left[Z_{0,a}Z_{0,a'}\psi_{-m}(v)\psi_{-l}(u)\right]\right|^2 dW(v)dW(u)\left[1+o(1)\right]
\end{aligned}
\tag{A.27}
$$

where we have used that fact that $E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m}(v)\psi_{q,-l}(u)\right] \to E\left[Z_{0,a}Z_{0,a'}\psi_{-m}(v)\psi_{-l}(u)\right]$ as $q\to\infty$ given Assumption A.2. Put $C(0,m,l) \equiv E\left[\left(Z_{0,a}Z_{0,a'} - \sigma(a,a')\right)\psi_{-m}(v)\psi_{-l}(u)\right]$. Then

$$
\begin{aligned}
E\left[Z_{0,a}Z_{0,a;}\psi_{-m}(v)\psi_{-l}(u)\right] &= C(0,m,l)+\sigma(a,a')\sigma_{l-m}(v,u) \\
\left|E\left[Z_{0,a}Z_{0,a'}\psi_{-m}(v)\psi_{-l}(u)\right]\right|^2 &= |C(0,m,l)|^2+\sigma(a,a')^2\left|\sigma_{l-m}(v,u)\right|^2+2\sigma(a,a')\operatorname{Re}\left[C(0,m,l)\sigma^*_{l-m}(v,u)\right]
\end{aligned}
$$

Given $\sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|C(0,m,l)|\le C$ and $|k(\cdot)|\le 1$, we have

$$
\begin{aligned}
& var\left(2\operatorname{Re}\widetilde{V}_q\right) \\
= & \sum\sum_{a,a'=i \text{ and } (i,j) \text{ with } i,j=1,\cdots d} 2\sigma(a,a')^2\sum_{m=1}^{q}\sum_{l=1}^{q}k^2(m/p)k^2(l/p)\int\int\left|\sigma_{l-m}(v,u)\right|^2dW(v)dW(u)\left[1+o(1)\right] \\
= & \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} 2\sigma(a,a')^2p\sum_{c=1-q}^{q-1}\left[p^{-1}\sum_{m=c+1}^{q}k^2(m/p)k^2\left[(m-c)/p\right]\right]\int\int\left|\sigma_c(v,u)\right|^2dW(v)dW(u)\left[1+o(1)\right] \\
= & \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} 2\sigma(a,a')^2p\int_0^{\infty}k^4(z)dz\sum_{c=-\infty}^{\infty}\int\int\left|\sigma_c(v,u)\right|^2dW(v)dW(u)\left[1+o(1)\right] \\
= & \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} 4\pi\sigma(a,a')^2p\int_0^{\infty}k^4(z)dz\int\int\int_{-\pi}^{\pi}\left|f(\omega,v,u)\right|^2d\omega dW(v)dW(u)\left[1+o(1)\right]
\end{aligned}
$$

where we used the fact that for many given c, $p^{-1}\sum_{m=c+1}^{q}k^2(m/p)k^2\left[(m-c)/p\right]\to\int_0^{\infty}k^4(z)dz$ as $p\to\infty$ and $q/p\to 0$.

I now verify condition (A.25). By $E\left|H_{m,.\tau-2q-1,a}(v)\right|^4\le C\tau^2$ for $1\le m\le q$ given the $m.d.s.$ property of $\{Z_{q,\tau},\mathcal{F}_{\tau-1}\}$ and Rosenthal's inequality for martingale(see Hall and Heyde, 1980, p.23),

$$
\begin{aligned}
E\left|V_{q,a}(\tau)\right|^4 &\le \left[\sum_{m=1}^{q}a_n(m)\int\left(E\left|Z_{q,\tau,a}\psi_{q,\tau-m}(v)H_{m,.\tau-2q-1,a}(v)\right|^4\right)^{\frac{1}{4}}dW(v)\right]^4 \\
&\le C\tau^2\left[\sum_{m=1}^{q}a_n(m)\right]^4=O(p^4\tau^2/n^4) \qquad \text{(A.28)}
\end{aligned}
$$

Then recall that $V_q(\tau)=\sum_{a=i \text{ and } (i,j) \text{ with } i,j=1,\cdots d}V_{q,a}$ and use Jensen's inequality, we have

$$
\begin{aligned}
E\left|V_q(\tau)\right|^4 &= E\left|\sum_{a=i \text{ and } (i,j),\ i,j=1,\cdots d}V_{q,a}\right|^4\le E\left[\sum_{a=i \text{ and } (i,j) \text{ with } i,j=1,\cdots d}\left|V_{q,a}(\tau)\right|\right]^4 \\
&= E\left\{\left(d'\right)^4\left[\sum_{a=i \text{ and } (i,j),\ i,j=1,\cdots d}\left|V_{q,a}(\tau)\right|\frac{1}{d'}\right]^4\right\} \\
&\le E\left\{\left(d'\right)^4\sum_{a=i \text{ and } (i,j),\ i,j=1,\cdots d}\left|V_{q,a}(\tau)\right|^4\frac{1}{d'}\right\}=\left(d'\right)^3\sum_{a=i \text{ and } (i,j),\ i,j=1,\cdots d}E\left|V_{q,a}(\tau)\right|^4=O(p^4\tau^2/n^4)
\end{aligned}
$$

where the last equality uses (A.28). It then follows that

$$\sum_{\tau=2q+2}^{n} E\left|V_q(\tau)\right|^4 = O(p^4/n) = o(p^2) \text{ given } p^2/n \to 0.$$

Therefore (A.25) is proved.

Next I turn to verify condition (A.26). Let $\sigma_{q,\tau}^2(a,a') \equiv E\left(Z_{q,\tau,a}Z_{q,\tau,a'}|\mathcal{F}_{\tau-1}\right)$. Then

$$\begin{aligned}
E\left[V_q^2(\tau)|\mathcal{F}_{\tau-1}\right] &= \sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d}\sum \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int \sigma_{q,\tau}^2(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u) \\
&\qquad H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)dW(v)dW(u) \\
&= \sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d}\sum \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int E\left[\sigma_{q,\tau}^2(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
&\qquad H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)dW(v)dW(u) \\
&+ \sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d}\sum \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int \widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)H_{m,.\tau-2q-1,a}(v) \\
&\qquad H_{l,.\tau-2q-1,a'}(u)dW(v)dW(u) \\
&\equiv S_{1q}(\tau) + V_{1q}(\tau), \text{ say} \qquad\qquad (A.29)
\end{aligned}$$

where

$$\widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u) \equiv \sigma_{q,\tau}^2(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u) - E\left[\sigma_{q,\tau}^2(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]$$

We further decompose

$$\begin{aligned}
S_{1q}(\tau) &= \sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d}\sum \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int E\left[\sigma_{q,\tau}^2(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
&\qquad E\left[H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right]dW(v)dW(u) \\
&+ \sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d}\sum \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int E\left[\sigma_{q,\tau}^2(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
&\qquad \left\{H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u) - E\left[H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right]\right\}dW(v)dW(u) \\
&\equiv E\left[V_q^2(\tau)\right] + S_{2q}(\tau), \text{ say} \qquad\qquad (A.30)
\end{aligned}$$

where

$$E\left[V_q^2(\tau)\right] \equiv \sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d}\sum \sum_{m=1}^{q}\sum_{l=1}^{q}(\tau-q-1)a_n(m)a_n(l)\int\int \left|E\left[\sigma_{q,\tau}^2(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]\right|dW(v)dW(u)$$

Put

$$Z_{q,s,aa'}^{m,l}(v,u) \equiv Z_{q,s,a}Z_{q,s,a'}\psi_{q,s-m}(v)\psi_{q,s-l}(u) - E\left[Z_{q,s,a}Z_{q,s,a'}\psi_{q,s-m}(v)\psi_{q,s-l}(u)\right]$$

Then write

$$
\begin{aligned}
S_{2q}(\tau) = & \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=1}^{\tau-2q-1} Z^{m,l}_{q,s,aa'}(v,u)dW(v)dW(u) \\
& + \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=2}^{\tau-2q-1} \sum_{c=1}^{s-1} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u) \\
\equiv & V_{2q}(\tau) + S_{3q}(\tau), \text{ say} \qquad \text{(A.31)}
\end{aligned}
$$

where

$$
\begin{aligned}
S_{3q}(\tau) = & \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=2}^{\tau-2q-1} \sum_{0<s-c\le 2q} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u) \\
& + \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=2}^{\tau-2q-1} \sum_{s-c>2q} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u) \\
\equiv & V_{3q}(\tau) + V_{4q}(\tau), \text{ say} \qquad \text{(A.32)}
\end{aligned}
$$

It follows from (A.29)-(A.32) that

$$\sum_{\tau=2q+2}^{n} \left\{E\left[V_q^2(\tau)|\mathcal{F}_{\tau-1}\right] - E\left[V_q^2(\tau)\right]\right\} = \sum_{h=1}^{4}\sum_{\tau=2q+2}^{n} V_{hq}(\tau)$$

Then it is sufficient to show Lemmas A.12-A.15 below, which imply that

$$E\left|\sum_{\tau=2q+2}^{n} \left\{E\left[V_q^2(\tau)|\mathcal{F}_{\tau-1}\right] - E\left[V_q^2(\tau)\right]\right\}\right|^2 = o(p^2)$$

given $q = p^{1+\frac{1}{4b-2}}\left(\ln^2 n\right)^{\frac{1}{2b-1}}$ and $p = cn^{\lambda}$ for $0 < \lambda < \left(3 + \frac{1}{4b-2}\right)^{-1}$. Thus condition (A.26) holds and so $M_{0q}(p) \to^d N(0,1)$ by Brown's theorem. ■

Lemma A.12: Let $V_{1q}(\tau)$ be defined as in (A.29). Then $E\left|\sum_{\tau=2q+2}^{n} V_{1q}(\tau)\right|^2 = O(qp^4/n)$.

Lemma A.13: Let $V_{2q}(\tau)$ be defined as in (A.31). Then $E\left|\sum_{\tau=2q+2}^{n} V_{2q}(\tau)\right|^2 = O(qp^4/n)$.

Lemma A.14: Let $V_{3q}(\tau)$ be defined as in (A.32). Then $E\left|\sum_{\tau=2q+2}^{n} V_{3q}(\tau)\right|^2 = O(qp^4/n)$.

Lemma A.15: Let $V_{4q}(\tau)$ be defined as in (A.32). Then $E\left|\sum_{\tau=2q+2}^{n} V_{4q}(\tau)\right|^2 = O(p)$.

Proof of Lemma A.12: Let

$$V_{1q,aa'}(\tau) \equiv \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int \widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)dW(v)dW(u)$$

Then from (A.29), $V_{1q}(\tau) = \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} V_{1q,aa'}(\tau)$. Recall the definition of $\widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)$ in (A.29). Noting that $\widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)$ is independent of $\left\{H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right\}$ and that $\widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)$ is independent of $\widetilde{Z}_{q,c,aa'}^{m,l}(v,u)$ for $\tau - c > 2q$ and $1 \le m,l \le q$, we have

$$\begin{aligned}
& E\left|\sum_{\tau=2q+2}^{n} \widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right|^2 \\
\le & \sum_{\tau=2q+2}^{n}\sum_{|\tau-c|\le 2q} E\left|\widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)\widetilde{Z}_{q,c,aa'}^{m,l}(v,u)\right|\left(E\left|H_{m,.\tau-2q-1,a}(v)\right|^4\right)^{\frac{1}{4}}\left(E\left|H_{l,.\tau-2q-1,a'}(u)\right|^4\right)^{\frac{1}{4}} \\
& \times\left(E\left|H_{m,.c-2q-1,a}(v)\right|^4\right)^{\frac{1}{4}}\left(E\left|H_{l,.c-2q-1,a'}(u)\right|^4\right)^{\frac{1}{4}} \\
= & O(n^3q)
\end{aligned}$$

where we have used the fact that $E\left|H_{m,.\tau-2q-1,a}(v)\right|^4 \le Cn^2$ for $1 \le m \le q$ and any a. It then follows by Minkowski's inequality and (A.11) that

$$\begin{aligned}
& E\left|\sum_{\tau=2q+2}^{n} V_{1q,aa'}(\tau)\right|^2 \\
\le & \left[\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\left(E\left|\sum_{\tau=2q+2}^{n}\int\int \widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right|^2\right)^{\frac{1}{2}}\right]^2 \\
= & O(qp^4/n) \qquad \text{(A.33)}
\end{aligned}$$

An application of Jensen's inequality implies that

$$E\left|\sum_{\tau=2q+2}^{n} V_{1q}(\tau)\right|^2 = E\left|\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d}\left[\sum_{\tau=2q+2}^{n} V_{1q,aa'}(\tau)\right]\right|^2$$
$$\leq d' \sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} E\left|\sum_{\tau=2q+2}^{n} V_{1q,aa'}(\tau)\right|^2 = O(qp^4/n) \qquad \text{(A.34)}$$

where (A.33) is used in the last equality. This completes the proof of Lemma A.12. ■

Proof of Lemma A.13: Let

$$V_{2q,aa'}(\tau) \equiv \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]\sum_{s=1}^{\tau-2q-1} Z_{q,s,aa'}^{m,l}(v,u)dW(v)dW(u)$$

Then from (A.31), $V_{2q}(\tau) = \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} V_{2q,aa'}(\tau)$. Recall the definition of $Z_{q,s,aa'}^{m,l}(v,u)$ in (A.31). Noting that $\left\{Z_{q,s,aa'}^{m,l}(v,u)\right\}_{m,l=1}^{q}$ is independent of $\left\{Z_{q,c,aa'}^{m,l}(v,u)\right\}_{m,l=1}^{q}$ for $|s-c| > 2q$ where q is sufficiently large, we have

$$E\left|\sum_{s=1}^{\tau-q-1} Z_{q,s,aa'}^{m,l}(v,u)\right|^2 = \sum_{s=1}^{\tau-q-1}\sum_{|s-c|\leq 2q} E\left[Z_{q,s,aa'}^{m,l}(v,u)Z_{q,c,aa'}^{m,l}(v,u)\right] \leq 2C\tau q \qquad \text{(A.35)}$$

It then follows that

$$E\left|\sum_{\tau=2q+2}^{n} V_{2q,aa'}(\tau)\right|^2 \leq \left\{\sum_{\tau=2q+2}^{n}\left[E\left|V_{2q,aa'}(\tau)\right|^2\right]^{\frac{1}{2}}\right\}^2$$
$$\leq \left\{\sum_{\tau=2q+2}^{n}\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\left|E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]\right|\right.$$
$$\left.\times\left(E\left|\sum_{s=1}^{\tau-2q-1} Z_{q,s,aa'}^{m,l}(v,u)\right|^2\right)^{\frac{1}{2}} dW(v)dW(u)\right\}^2$$
$$= O(qp^4/n) \qquad \text{(A.36)}$$

where we have used (A.35). Then the same reasoning as that of (A.34) which uses Jensen's inequality and (A.36) gives us the desired result ■

Proof of Lemma A.14: Let

$$
\begin{aligned}
& V_{3q,aa'}(\tau) \\
\equiv\ & \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=2}^{\tau-2q-1}\sum_{0<s-c\leq 2q} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u)
\end{aligned}
$$

Then from (A.32), $V_{3q}(\tau) = \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} V_{3q,aa'}(\tau)$. By the same reasoning as that for the proof of Lemma A.13, it is sufficient to show that $E\left|\sum_{\tau=2q+2}^{n} V_{3q,aa'}(\tau)\right|^2 = O(qp^4/n)$. This follows from Minkowski's inequality and

$$
\begin{aligned}
& E\left|V_{3q,aa'}(\tau)\right|^2 \\
\leq\ & \left\{\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int \left|E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]\right|\right. \\
& \left.\times \left(\sum_{s=1}^{\tau-2q-1} E\left|Z_{q,s,a}\psi_{q,s-m}(v)\sum_{s-c\leq 2q} Z_{q,c,a'}\psi_{q,c-l}(u)\right|^2\right)^{\frac{1}{2}} dW(v)dW(u)\right\}^2 \\
& \leq 2C\tau q\left[\sum_{m=1}^{q} a_n(m)\right]^4 = O(\tau qp^4/n) \qquad \blacksquare
\end{aligned}
$$

Proof of Lemma A.15: Let

$$
\begin{aligned}
& V_{4q,aa'}(\tau) \\
\equiv\ & \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=2}^{\tau-2q-1}\sum_{s-c>2q} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u)
\end{aligned}
$$

Then from (A.32), $V_{4q}(\tau) = \sum\sum_{a,a'=i \text{ and } (i,j),\ i,j=1,\cdots d} V_{4q,aa'}(\tau)$. By the same reasoning as that for the proof of Lemma A.13, it is sufficient to show that $E\left|\sum_{\tau=2q+2}^{n} V_{4q,aa'}(\tau)\right|^2 = O(p)$. This follows from Minkowski's inequality, $p \to \infty$ and

$$
\begin{aligned}
& E\left|V_{4q,aa'}(\tau)\right|^2 \\
\leq\ & E\Bigg|\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \qquad\qquad \times \sum_{s=2q+2}^{\tau-2q-1} Z_{q,s,a}\psi_{q,s-m}(v)\sum_{c=1}^{s-2q-1} Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u)\Bigg|^2 \\
\leq\ & \sum_{m_1=1}^{q}\sum_{m_2=1}^{q}\sum_{l_1=1}^{q}\sum_{l_2=1}^{q} a_n(m_1)a_n(m_2)a_n(l_1)a_n(l_2)\int\int\int\int E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m_1}(v_1)\psi_{q,-l_1}(v_2)\right] \\
& \times E\left[Z_{q,0,a}Z_{q,0,a'}\psi^*_{q,-m_2}(u_1)\psi^*_{q,-l_2}(u_2)\right]\sum_{s=2q+2}^{\tau-2q-1} E\left[Z_{q,s,a}Z_{q,s,a'}\psi_{q,s-m_1}(v_1)\psi_{q,s-m_2}(u_2)\right] \\
& \times \sum_{c=1}^{s-2q-1} E\left[Z_{q,c,a}Z_{q,c,a'}\psi^*_{q,c-l_1}(v_2)\psi^*_{q,c-l_2}(u_2)\right]dW(v_1)dW(v_2)dW(u_1)dW(u_2) \\
=\ & O(\tau^2 p/n^4)
\end{aligned}
$$

by Assumption A.2 and A.8(i.e., $\sum_{m=1}^{\infty}|\sigma_m(v,u)|\leq C$, $\sum_{m=1}^{\infty}\sum_{l=1}^{\infty}E\left|\left(Z_{0,a}Z_{0,a'}-\sigma(a,a')\right)\psi_{-m}(v)\psi_{-l}(u)\right|$). ■

Proof of Theorem4. It is sufficient to prove the following Theorems A4 and A5. ■

TheoremA4. Under the conditions of Theorem4, $(p^{\frac{1}{2}}n)\left[\widehat{M}_0(p)-M_0(p)\right]\to^p 0$.

TheoremA5. Under the conditions of Theorem4 and for $a=i$ and ij, $i,j=1,\cdots d$,

$$
(p^{\frac{1}{2}}/n)M_0(p)\to^p\left[2D\int_0^{\infty}k^4(z)dz\right]^{-1/2}\int\int_{-\pi}^{\pi}\left|f_a^{(0,1,0)}(\omega,0,v)-f_{0,a}^{(0,1,0)}(\omega,0,v)\right|^2 d\omega dW(v) \tag{A.37}
$$

and therefore

$$
(p^{\frac{1}{2}}/n)M_0(p)\to^p\left[2D\int_0^{\infty}k^4(z)dz\right]^{-1/2}\int\int_{-\pi}^{\pi}\left\|f^{(0,1,0)}(\omega,0,v)-f_0^{(0,1,0)}(\omega,0,v)\right\|^2 d\omega dW(v)
$$

Proof of Theorem A4. It suffices to show that

$$
n^{-1}\int\sum_{m=1}^{n}k^2(m/p)n_m\left[\left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2-\left\|\widetilde{\sigma}_m^{(1,0)}(0,v)\right\|^2\right]dW(v)\to^p 0 \tag{A.38}
$$

$p^{-1}\left[\widehat{C}_0(p)-\widetilde{C}_0(p)\right]=O_p(1)$, and $p^{-1}\left[\widehat{D}_0(p)-\widetilde{D}_0(p)\right]=o_p(1)$, where $\widetilde{C}_0(p)$ and $\widetilde{D}_0(p)$ are defined in the same way as $\widehat{C}_0(p)$ and $\widehat{D}_0(p)$ in (3.5) with $\{Z_\tau\}_{\tau=1}^n$ replacing $\{\widehat{Z}_\tau\}_{\tau=1}^n$. I focus on the proof of (A.38) to save space; the proofs for $p^{-1}\left[\widehat{C}_0(p)-\widetilde{C}_0(p)\right]=O_p(1)$, and $p^{-1}\left[\widehat{D}_0(p)-\widetilde{D}_0(p)\right]=o_p(1)$ are straightforward. Because (A.5) implies that

$$\begin{aligned} & n^{-1}\int\sum_{m=1}^{n}k^2(m/p)n_m\left[\left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2-\left\|\widetilde{\sigma}_m^{(1,0)}(0,v)\right\|^2\right]dW(v) \\ = & \sum_{a=i \text{ and } (i,j),\ i,j=1,\cdots d} n^{-1}\int\sum_{m=1}^{n}k^2(m/p)n_m\left[\left|\widehat{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2-\left|\widetilde{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2\right]dW(v), \end{aligned}$$

then it suffices to prove that

$$n^{-1}\int\sum_{m=1}^{n}k^2(m/p)n_m\left[\left|\widehat{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2-\left|\widetilde{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2\right]dW(v)\to^p 0,\text{ for } a=i \text{ and } (i,j),\ i,j=1,\cdots d \quad \text{(A.39)}$$

We shall show this only for the case $a=i$ with $i=1,\cdots d$; the proofs for all other cases are similar.

Since the proof of Theorem A.5 does not depend on Theorem A4, it follows from (A.37) that

$$n^{-1}\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2dW(v)=O_p(1)\text{ for } i=1,\cdots d \quad \text{(A.40)}$$

By (A.11), (A.40), (A.8) and Cauchy-Schwartz inequality, it is sufficient for (A.39) to prove that $n^{-1}\widehat{A}_1=o_p(1)$, where $\widehat{A}_1$ is defined as in (A.8). Then (A.9) implies further that it is enough to show

$$n^{-1}\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left|\widehat{B}_{hm}(v)\right|^2dW(v)=o_p(1),\text{ for } h=1,\cdots,6$$

I first consider the case $h=1$. By Cauchy-Schwartz inequality and $\left|\widehat{\delta}_\tau(v)\right|\leqq 2$,

$$\left|\widehat{B}_{1m}(v)\right|^2\leqq\left[n_m^{-1}\sum_{\tau=m+1}^{n}\left(\widehat{Z}_{\tau,i}-Z_{\tau,i}\right)^2\right]\left[n_m^{-1}\sum_{\tau=m+1}^{n}\left|\widehat{\delta}_\tau(v)\right|^2\right]\leqq n_m^{-1}\sum_{\tau=1}^{n}\left(\widehat{Z}_{\tau,i}-Z_{\tau,i}\right)^2$$

Then it follows from (A.3) and (A.11) and Assumption A.6 that

$$n^{-1}\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left|\widehat{B}_{1m}(v)\right|^2dW(v)\leqq\left[\sum_{\tau=1}^{n}\left(\widehat{Z}_{\tau,i}-Z_{\tau,i}\right)^2\right]\sum_{\tau=m+1}^{n}a_n(m)\left[\int dW(v)\right]^2=O_p(p/n)$$

The proof for case $h=2$ is similar, noting that

$$\left|n_m^{-1}\sum_{\tau=m+1}^{n}\left(\widehat{Z}_{\tau,i}-Z_{\tau,i}\right)\right|^2\leqq n_m^{-1}\sum_{\tau=m+1}^{n}\left|\widehat{Z}_{\tau,i}-Z_{\tau,i}\right|^2$$

Next consider the case $h=3$. Still by the Cauchy-Schwartz inequality, I have

$$\left|\widehat{B}_{3m}(v)\right|^2 \leqq \left(n_m^{-1}\sum_{\tau=1}^{n} Z_{\tau,i}^2\right) n_m^{-1} \sum_{\tau=m+1}^{n} \left|\widehat{\delta}_{\tau-m}(v)\right|^2 \leqq \|v\|^2 \left(n_m^{-1}\sum_{\tau=1}^{n} Z_{\tau,i}^2\right) n_m^{-1} \sum_{\tau=m+1}^{n} \left(\widehat{Z}_{\tau,i} - Z_{\tau,i}\right)^2$$

It then follows that

$$\begin{aligned}
& n^{-1}\int \sum_{m=1}^{n-1} k^2(m/p) n_m \left|\widehat{B}_{3m}(v)\right|^2 dW(v) \\
\leqq \quad & \left(n^{-1}\sum_{\tau=1}^{n} Z_{\tau,i}^2\right)\left[n^{-1}\sum_{\tau=1}^{n}\left(\widehat{Z}_{\tau,i} - Z_{\tau,i}\right)^2\right]\sum_{\tau=1}^{n-1} k^2(m/p)\int \|v\|^2 dW(v) \\
= \quad & O_p(p/n)
\end{aligned}$$

The proof for the cases $h = 4, 5, 6$ is similar to the case $h = 3$, noting that

$$\left|n_m^{-1} \sum_{\tau=m+1}^{n} \widehat{\delta}_{\tau}(v)\right|^2 \leqq n_m^{-1} \sum_{\tau=m+1}^{n} \left|\widehat{\delta}_{\tau}(v)\right|^2$$

This completes the proof of Theorem A4. ■

Proof of Theorem A.5. The proof is a straightforward extension for that of Hong(1999, Proof of Theorem 5), for the case $(m, l) = (1, 0)$ and $W_1(\cdot) = \delta(\cdot)$, the Dirac delta function. I omit it here to save space. Note that Assumption A.8 is needed here. ■

Proof of Theorem5. It is sufficient to prove the Theorems A6 and A7 below. ■

TheoremA6. Under the conditions of Theorem5, $\widehat{M_0}(\widehat{p}) - M_0(\widehat{p}) = o_p(1)$.

TheoremA7. Under the conditions of Theorem5, $M_0(\widehat{p}) - M_0(p) = o_p(1)$.

Proof of TheoremA6. Define

$$\widehat{B} \equiv \sum_{m=1}^{n-1} k^2(m/\widehat{p}) n_m \int \left[\left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2 - \left\|\widetilde{\sigma}_m^{(1,0)}(0,v)\right\|^2\right] dW(v)$$

Then it suffices to show that $p^{-\frac{1}{2}}\widehat{B} = o_p(1)$, $p^{-\frac{1}{2}}\left[\widehat{C}_0(\widehat{p}) - \widetilde{C}_0(\widehat{p})\right] = o_p(1)$, and $p^{-1}\left[\widehat{D}_0(\widehat{p}) - \widetilde{D}_0(\widehat{p})\right] = o_p(1)$. I only show $p^{-\frac{1}{2}}\widehat{B} = o_p(1)$ here to save space; the proof of the other two is similar. Since

$$
\begin{aligned}
\widehat{B} &= \sum_{a=i \text{ and } (i,j),\ i,j=1,\cdots d} \left\{ \sum_{m=1}^{n-1} k^2(m/\widehat{p}) n_m \int \left[\left| \widehat{\sigma}_{m,a}^{(1,0)}(0,v) \right|^2 - \left| \widetilde{\sigma}_{m,a}^{(1,0)}(0,v) \right|^2 \right] dW(v) \right\} \\
&\equiv \sum_{a=i \text{ and } (i,j),\ i,j=1,\cdots d} \widehat{B}_a
\end{aligned}
$$

then it is sufficient to show $p^{-\frac{1}{2}} \widehat{B}_a = o_p(1)$ for $a = i$ and ij, $i,j = 1, \cdots d$. We shall only show this holds for $a = i$; the proofs for all other cases are similar.

By the conditions on $k(\cdot)$ implied by Assumption A.5, there exists a symmetric monotonic decreasing function $k_0(z)$ for $z > 0$ such that $|k(z)| \leqq k_0(z)$ for all $z > 0$ and $k_0(\cdot)$ satisfies Assumption A.5. Then for any constants $\epsilon, \eta > 0$,

$$
P\left(p^{-\frac{1}{2}} \left| \widehat{B}_i \right| > \epsilon \right) \leq P\left(p^{-\frac{1}{2}} \left| \widehat{B}_i \right| > \epsilon, |\widehat{p}/p - 1| \leq \eta \right) + P\left(|\widehat{p}/p - 1| > \eta \right)
$$

where the second term vanishes asymptotically for all $\eta > 0$ and given $\widehat{p}/p - 1 \to^p 0$. Therefore, by the definition of convergence in probability, it remains to show that the first terms also vanishes as $n \to \infty$.

Because $|\widehat{p}/p - 1| \leq \eta$ implies $\widehat{p} \leq (1+\eta)p$, for $|\widehat{p}/p - 1| \leq \eta$

$$
p^{-\frac{1}{2}} \left| \widehat{B}_i \right| \leq (1+\eta)^{\frac{1}{2}} \left[(1+\eta)p \right]^{-\frac{1}{2}} \sum_{m=1}^{n-1} k_0^2 \left[m/(1+\eta)p \right] n_m \left[\left| \widehat{\sigma}_{m,i}^{(1,0)}(0,v) \right|^2 - \left| \widetilde{\sigma}_{m,i}^{(1,0)}(0,v) \right|^2 \right] \to^p 0
$$

for any $\eta > 0$ given (A.9), where the inequality follows from $|k(z)| \leqq k_0(z)$ for all $z > 0$. This completes the proof of Theorem A6. ■

Proof of Theorem A7. The proof is a straightforward extension from those of Theorem A.7 of Hong and Lee(2005) which follows a reasoning analogous to the proof of Hong(1999, Theorem4). Note that the latter uses an assumption which is exactly the same as Assumption A.10. That is, Assumption A.10 is also used in this proof. ■

References

[1] Ahn, D., and B. Gao, 1999, “A Parametric Nonlinear Model of Term Structure Dynamics,” Review of Financial Studies, 12, 721–762.

[2] Ait-Sahalia, Y. 1996a, "Testing Continuous-Time Models of the Spot Interest Rate," Review of Financial Studies, 9, 385-426.

[3] Ait-Sahalia, Y. 1996b, "Nonparametric Pricing of Interest Rate Derivative Securities," Econometrica, 64, 527-560.

[4] Ait-Sahalia, Y. 1997, "Do Interest Rates Really Follow Continuous-Time Markov Diffusions?" working paper, Princeton University.

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

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

[7] Ait-Sahalia, Y., J. Fan and H. Peng (2008): "Nonparametric Transition-Based Tests for Diffusions" forthcoming in Annals of Statistics.

[8] Ait-Sahalia Y., Hansen, L. and Scheinkman, J. 2004 "Operator Methods for Continuous-Time Markov Processes", Handbook of Financial Econometrics.

[9] Ait-Sahalia, Y. and Jacod, J. (2008) "Testing for Jumps in a Discretely Observed Process," forthcoming in Annals of Statistics.

[10] Ait-Sahalia, Y., Mykland, P. and Zhang, L. 2005, "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data", Journal of the American Statistical Association, 100, 1394-1411.

[11] Andersen, T.G., Benzoni, L. and Lund, J. 2002, "An Empirical Investigation of Continuous-Time Models for Equity Returns," Journal of Finance 57 (2002): 1239-1284.

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

[13] Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

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

[15] Bandi, F. M. and Phillips, P.C.B. (2007), "A Simple Approach to the Parametric Estimation of Potentially Nonstationary Diffusions.", Journal of Econometrics, 137(2), pp. 354-95.

[16] Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, 1–37.

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

[18] Bierens, H. 1982: "Consistent Model Specification Tests," Journal of Econometrics, 20, 105-134.

[19] Bierens, H., and Ploberger, W. 1997 "Asymptotic Theory of Integrated Conditional Moments Tests", Econometrica, 58, 1129-1151.

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

[21] Brillinger, D.R. and Rosenblatt, M. 1967a. Asymptotic theory of estimates of kth order spectra. In Spectral Analysis of Time Series, ed. B. Harris. New York: Wiley.

[22] Brillinger, D.R. and Rosenblatt, M. 1967b. Computation and interpretation of the kth order spectra. In Spectral Analysis of Time Series, ed. B. Harris. New York: Wiley.

[23] Brown, B.M. 1971, "Martingale Limit Theorems," Annals of Mathematical Statistics, 42, 59–66.

[24] Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders, 1992, "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," Journal of Finance, 47, 1209–1227.

[25] Chapman, D., and N. Pearson, 2000, "Is the Short Rate Drift Actually Nonlinear?" Journal of Finance, 55, 355–388.

[26] Chen, S. X., Gao, J. and Tang, C. (2008). A test for model specification of diffusion processes. The Annals of Statistics 36, 167-198.

[27] Chen, B. and Hong, Y. 2008a, "Characteristic function-based testing for multifactor continuous-time models via nonparametric regression", working paper.

[28] Chen, B. and Hong, Y. 2008b, "Testing for the Markov property in time series", working paper.

[29] Conley, T.G., L.P. Hansen, E.G.J. Luttmer, and J.A. Scheinkman, 1997, "Short-Term Interest Rates as Subordinated Diffusions," Review of Financial Studies, 10, 525–578.

[30] Corradi, V. and N.R. Swanson. 2005. "Bootstrap specification tests for diffusion process", Journal of Econometrics 124, 117-148.

[31] Corradi, V. and H. White. 1999. "Specification Tests for the Variance of a Diffusion." Journal of Time Series Analysis 20: 253-270.

[32] Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," Econometrica, 53, 385–407.

[33] Duffie, D., L. Pedersen, and K. Singleton, 2003, "Modeling Credit Spreads on Sovereign Debt: A Case Study of Russian Bonds," Journal of Finance, 55, 119–159.

[34] Dynkin, E.B. (1965) Markov Processes, Volume I, Springer-Verlag. 32

[35] Fan, Y. and Fan, J., 2008, "Testing and detecting jumps based on a discretely observed process", Manuscript.

[36] Fan, J.and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-Frequency financial data,Jour. Ameri. Statist. Assoc.,102,1349-1362.

[37] Fan, J. & Zhang, C. 2003, "A re-examination of Stanton's diffusion estimation with applications to financial model validation", Journal of the American Statistical Association 457, 118–134.

[38] Gallant, A.R., and G. Tauchen, 1996, "Which Moments to Match?" Econometric Theory, 12, 657–681.

[39] Gao, J. and Casas, I. 2008, "Specification testing in discretized diffusion models". Journal of Econometrics 147, 131-140.

[40] Hall, P., and C. Heyde, 1980, Martingale Limit Theory and Its Application, Academic Press, New York.

[41] Hannan, E. 1970, Multiple Time Series, Wiley, New York.

[42] Hansen, L.P., and J.A. Scheinkman, 1995, “Back to the Future: Generating Moment Implications for Continuous Time Markov Processes,” Econometrica, 63, 767–804.

[43] Hansen, L.P., and J.A. Scheinkman, 2003, "Semigroup pricing", manuscript.

[44] Hansen, L.P., J.A. Scheinkman, and Nizar Touzi. 1998. "Spectral methods for identifying scalar diffusions". Journal of Econometrics, Volume 86, Issue 1, September, Pages 1-32

[45] Hong, Y. (1999): "Hypothesis Testing in Time Series via the Empirical Characteristic Function: a Generalized Spectral Density Approach," Journal of the American Statistical Association, 94, 1201-1220

[46] Hong, Y. and Lee, Y. 2004, "Specification testing for multivariate time series volatility models", working paper.

[47] Hong, Y. and Lee, Y. 2005, “Generalized Spectral Tests for Conditional Mean Specification in Time Series with Conditional Heteroskedasticity of Unknown Form”, Review of Economic Studies, Vol. 72, 499-541

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

[49] Hong, Y. and Song, Z. 2008a, "Generalized bispectrum: a new measure of serial dependence in time series", working paper, Cornell University.

[50] Hong, Y. and Song, Z. 2008b, "Hypothesis testing of joint serial dependence in time series via generalized bispectrum", working paper, Cornell University.

[51] Jiang, G., and J. Knight, 1997, “A Nonparametric Approach to the Estimation of Diffusion Processes with an Application to a Short-Term Interest Rate Model,” Econometric Theory, 13, 615–645.

[52] Jiang, G., and J. Knight, 2002, “Estimation of Continuous-Time Processes via the Empirical Characteristic Function,” Journal of Business and Economic Statistics, 20, 198–212.

[53] Johannes, M., 2004, “The Statistical and Economic Role of Jumps in Interest Rates,” Journal of Finance, 59, 227–260.

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

[55] Kanaya, S. 2007, "Non-Parametric Specification Testing for Continuous-Time Markov Processes:Do the Processes Follow Diffusions?", Manuscript, University of Wisconsin.

[56] Karatzas, I. und St. E. Shreve, 1988. "Brownian motion and stochastic calculus". Graduate Texts in Mathematics, 113 Springer-Verlag, New York.

[57] Kessler, M., Sorensen, M., 1996. Estimating equations based on eigenfunctions for a discretely observed diffusion process.

[58] Kristensen, D. 2008a, "Nonparametric Estimation and Misspecification Testing of Diffusion Models", working paper, Columbia University.

[59] Kristensen, D. 2008b, "Pseudo-Maximum-Likelihood Estimation in Two Classes of Semiparametric Diffusion Models", working paper, Columbia University.

[60] Lee, S. and Mykland, P. 2008, "Jumps in financial markets: a new nonparametric test and jump dynamics", Review of financial studies

[61] Leonov, V.P. and Shiryaev, A.N. 1959. "On a method of calculation of semi-invariants", (English translation). Theor. Prob. Appl. 4 319-329.

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

[63] Lo, A.W., 1988, "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," Econometric Theory, 4, 231–247.

[64] Lobato, I. 2002, "A consistent test for the martingale difference hypothesis", Econometric Reviews.

[65] Nelson D.B. (1990), "ARCH Models as Diffusion Approximations", Journal of Econometrics, 45, 7–38.

[66] Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time series study. Journal of Financial Economics 63, 3–50.

[67] Park, J. 2008, "Martingale regression and time change", working paper.

[68] Park, J. and Whang, Y.J. 2003, "Testing for the martingale hypothesis", working paper, Department of Economics, Rice University and Department of Economics, Korea University.

[69] Pedersen, A. R., 1995, "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," Scandinavian Journal of Statistics, 22, 55–71.

[70] Priestley, M.B. 1981. Spectral Analysis and Time Series. London: Academic Press.

[71] Pritsker, M., 1998, "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models," Review of Financial Studies, 11, 449–487.

[72] Protter, P. 2005 "Stochastic Integration and Differential Equations, Second Edition", Version 2. Springer-Verlag, Heidelberg.

[73] Revuz, D and Yor, M. 2005 "Continuous martingales and Brownian motion", volume 293 of Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, Berlin.

[74] Rogers, L.C.G., and Williams, D. (2000) Diffusions, Markov Processes and Martingales, Vol. 1, 2nd edition, Cambridge University Press.

[75] Schoutens, W. 2003. "Levy Processes in Finance". Wiley, New York.

[76] Singleton, K., 2001, “Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function,” Journal of Econometrics, 102, 111–141.

[77] Stanton, R., 1997, “A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk,” Journal of Finance, 52, 1973–2002.

[78] Sundaresan, S., 2001, “Continuous-Time Methods in Finance: A Review and an Assessment,” Journal of Finance, 55, 1569–1622.

[79] Stinchcombe M. and H. White (1998): "Consistent Specification Testing with Nuisance Parameters Present only Under the Alternative," Econometric Theory, 14, 295-324.

[80] Stratonovich, R.L., 1963. "Topics in the Theory of Random Noise", 1, (English translation). Gordon and Breach, New York.

[81] Stroock, D.W. and Varadhan, S.R.S. 1969, "Diffusion processes with continuous coefficients I & II", Comm. Pure & Appl. Math. 22, 345-400 and 479-530.

[82] Tauchen, G., 1997, “New Minimum Chi-Square Methods in Empirical Finance,” in D. Kreps and K.Wallis (eds.), Advances in Econometrics: Seventh World Congress, Cambridge University Press, Cambridge, UK.

[83] Vasicek, O., 1977, “An Equilibrium Characterization of the Term Structure,” Journal of Financial Economics, 5, 177–188.

[84] Zähle, H. 2008, "Weak approximations of SDEs by discrete-time processes", Journal of Applied Mathematics and Stochastic Analysis.

TABLE I

	Empirical Level Under Vasicek Model							
	$n = 250$		$n = 500$		$n = 1000$		$n = 1500$	
	10%	5%	10%	5%	10%	5%	10%	5%
	High Persistence							
5	0.313	0.298	0.163	0.152	0.165	0.143	0.114	0.085
10	0.362	0.350	0.202	0.184	0.142	0.132	0.114	0.085
15	0.372	0.341	0.187	0.175	0.151	0.138	0.114	0.085
20	0.325	0.274	0.168	0.146	0.162	0.145	0.114	0.085
	Low Persistence							
5	0.343	0.312	0.162	0.155	0.166	0.145	0.092	0.074
10	0.380	0.373	0.210	0.189	0.145	0.133	0.092	0.074
15	0.396	0.356	0.193	0.180	0.150	0.134	0.092	0.074
20	0.346	0.303	0.170	0.152	0.167	0.150	0.092	0.074

Notes : (i) 1000 iterations;

(ii): High persistent Vasicek model is model (6.1) with parameter values $\left(\kappa, \alpha, \sigma^2\right) = (0.214592, 0.089102, 0.000546)$; Low persistent Vasicek model is model (6.1) with parameter values $\left(\kappa, \alpha, \sigma^2\right) = (0.85837, 0.089102, 0.002185)$.

(iii):$\overline{p}$, the preliminary bandwidth, is used in a plug-in method to choose a data-dependent bandwidth $\widehat{p}_0$ with the Bartlett kernel used. The Bartlett kernel is also used for computing $\widehat{M_0}(\widehat{p}_0)$.

TABLE II

	Empirical Power Under DGP1-4					
	$n = 250$		$n = 500$		$n = 1000$	
	10%	5%	10%	5%	10%	5%
	CIR					
5	0.656	0.634	0.778	0.764	1.000	1.000
10	0.693	0.662	0.773	0.760	0.987	0.984
15	0.722	0.679	0.765	0.737	0.964	0.960
20	0.683	0.646	0.784	0.773	1.000	1.000
	CKLS					
5	0.684	0.659	0.798	0.763	1.000	1.000
10	0.690	0.667	0.782	0.754	1.000	1.000
15	0.716	0.701	0.764	0.752	1.000	1.000
20	0.689	0.673	0.758	0.747	1.000	1.000
	Ahn&Gao					
5	0.268	0.262	0.638	0.588	0.965	0.954
10	0.252	0.247	0.635	0.583	0.967	0.962
15	0.244	0.240	0.609	0.552	0.978	0.970
20	0.237	0.235	0.627	0.573	0.975	0.968
	$Ait - Sahalia$					
5	0.475	0.425	0.838	0.802	0.967	0.962
10	0.454	0.414	0.835	0.798	0.978	0.973
15	0.435	0.422	0.852	0.814	0.962	0.958
20	0.449	0.425	0.867	0.832	0.957	0.942

Notes : (i) 1000 iterations;

(ii) DGP1-4 are CIR model, Ahn and Gao's(1997) inverse-feller model, CKLS model and Ait-Sahalia's(1996) nonlinear drift model, given in equations (6.2)-(6.4).

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