Econometric Theory, **26**, 2010, 1115–1179.
doi:10.1017/S026646660999048X

CHARACTERISTIC FUNCTION–BASED TESTING FOR MULTIFACTOR CONTINUOUS-TIME MARKOV MODELS VIA NONPARAMETRIC REGRESSION

Bin Chen
University of Rochester

Yongmiao Hong
Cornell University
and
Xiamen University

We develop a nonparametric regression-based goodness-of-fit test for multifactor continuous-time Markov models using the conditional characteristic function, which often has a convenient closed form or can be approximated accurately for many popular continuous-time Markov models in economics and finance. An omnibus test fully utilizes the information in the joint conditional distribution of the underlying processes and hence has power against a vast class of continuous-time alternatives in the multifactor framework. A class of easy-to-interpret diagnostic procedures is also proposed to gauge possible sources of model misspecification. All the proposed test statistics have a convenient asymptotic $N(0,1)$ distribution under correct model specification, and all asymptotic results allow for some data-dependent bandwidth. Simulations show that in finite samples, our tests have reasonable size, thanks to the dimension reduction in nonparametric regression, and good power against a variety of alternatives, including misspecifications in the joint dynamics, but the dynamics of each individual component is correctly specified. This feature is not attainable by some existing tests. A parametric bootstrap improves the finite-sample performance of proposed tests but with a higher computational cost.

We thank the co-editor, Oliver Linton, and two referees for careful and constructive comments. We also thank Jianqing Fan, Robert Jarrow, Joon Park, Peter Phillips, George Tauchen, Dag Tjøstheim, Qiwei Yao, and seminar participants at the 2006 Symposium on Financial Engineering and Risk Management in Beijing, the 2006 International Symposium on Econometric Theory and Applications (SETA) in Xiamen, the North American Econometric Society Summer Meeting in Minneapolis, the 2006 Far Eastern Econometric Society Meeting in Beijing, and the 2008 International Symposium on Recent Developments of Time Series Econometrics in Xiamen for their comments. All remaining errors are solely ours. Hong acknowledges support from the National Science Foundation of China and the Cheung Kong Visiting Scholarship by the Chinese Ministry of Education and Xiamen University, China. Address correspondence to Bin Chen, Department of Economics, University of Rochester, Rochester, NY 14627, USA; e-mail: bchen8@mail.rochester.edu; Yongmiao Hong, Departments of Economics and Statistical Science, Cornell University, Ithaca, NY 14850, USA; e-mail: yh20@cornell.edu.

 0266-4666/10 $15.00

1. INTRODUCTION

Continuous-time Markov models are powerful analytic tools in modern finance and economics. Itô processes have been popularly adapted, and the more general Lévy processes have been the object of recent research for derivatives pricing in the literature (e.g., Carr and Wu, 2003, 2004; Chernov, Gallant, Ghysels, and Tauchen, 1999). There are several reasons for the popularity of continuous-time Markov models in finance and economics. First, continuous information flows into financial markets provide a justification for using continuous-time models, and the development of stochastic calculus provides a powerful tool for elegant mathematical treatment of continuous-time models. Second, the Markov assumption, which is a maintained condition for almost all continuous-time models in finance and economics, simplifies greatly the involved mathematical derivation. Under the Markov assumption, the conditional probability distribution of future values of the underlying process, conditional on the currently available information, depends only on the current value of the process, and the inclusion of any additional information available at the current time will not alter this conditional probability distribution. From an economic point of view, economic agents' rationality provides a solid justification for the Markov assumption. Economic agents update beliefs and make decisions sequentially. Their subjective beliefs about future uncertainty and optimal decision rules are often assumed to depend on the past information only via the current state.[1]

Econometric analysis of continuous-time models is generally more challenging than that of discrete-time dynamic models. Much progress has been made in the literature in estimating continuous-time models. For example, the Aït-Sahalia (2002) approximated maximum likelihood estimator (MLE), the Bates (2007) filtration-based MLE, the Chib, Pitt, and Shephard (2004) Markov chain Monte Carlo (MCMC) method, the Gallant and Tauchen (1996) efficient method of moments (EMM) method, and the Singleton (2001) conditional characteristic function–based maximum likelihood estimation (MLE-CCF) and conditional characteristic function–based general method of moments (GMM-CCF) methods have been proposed.[2] In contrast, there has been relatively little effort devoted to specification analysis and validation of continuous-time models. In a continuous-time framework, model misspecification generally renders inconsistent parameter estimators and their conventional variance-covariance matrix estimators, which could result in misleading conclusions on statistical inference. The validity of economic interpretations for model parameters also crucially depends on correct model specification. More importantly, a misspecified model can yield large errors in pricing, hedging, and risk management.

Nevertheless, economic theories usually do not suggest a concrete functional form for continuous-time Markov models. The choice of a model is somewhat arbitrary, often based on convenience and empirical experience of the practitioner. For example, in the pricing and hedging literature, a continuous-time model is often assumed to have a functional form that yields a closed-form pricing formula, as in the case of multivariate affine term structure models (ATSMs) (Dai and Singleton, 2000; Duffie and Kan, 1996). It is important to develop a reliable omnibus specification test for popular continuous-time Markov models. In addition, diagnostic procedures that focus on misspecification in certain directions (e.g., conditional mean, conditional variance, and conditional correlation) will be also useful for guiding further improvement of the model.

There has been some work on testing continuous-time models. Aït-Sahalia (1996a) develops a nonparametric test for univariate diffusion models. Aït-Sahalia (1996a) checks the adequacy of a diffusion model by comparing the model-implied stationary density with a smoothed kernel density estimator based on discretely sampled data.[3] Gao and King (2004) develop a simulation procedure to improve the finite-sample performance of the Aït-Sahalia (1996a) test. These tests are convenient to implement, but they may overlook a misspecified model with a correct stationary density.

Hong and Li (2005) develop a specification test for continuous-time models using the transition density, which can capture the full dynamics of a continuous-time process. Observing the fact that when a continuous-time model is correctly specified, the probability integral transform (PIT) of the observed sample with respect to the model-implied transition density is independent and identically distribution (i.i.d.) $U[0, 1]$, they check the joint hypothesis of i.i.d. $U[0, 1]$ using a nonparametric density estimator. The most appealing feature of this test is its robustness to persistent dependence in data because the PIT series is always i.i.d. $U[0, 1]$ under correct model specification. This approach, however, cannot be extended to a multivariate joint transition density, because it is well known that the PIT series with respect to a multivariate joint transition density is no longer i.i.d $U[0, 1]$ even if the model is correctly specified. Hong and Li (2005) apply their test to evaluate multivariate continuous-time models by considering the PIT for each individual state variable, with a suitable partitioning. This practice is valid, but it may fail to detect model misspecification in the joint dynamics of state variables.

Alternative tests for univariate diffusion models have recently been suggested in the literature. Aït-Sahalia, Fan, and Peng (2009) propose new tests by comparing the model-implied transition density and distribution functions with their nonparametric counterparts, respectively. Chen, Gao, and Tang (2008) develop a transition density–based test using a nonparametric empirical likelihood approach. Li (2007) tests the parametric specification of the diffusion function by measuring the distance between the model-implied diffusion function and its kernel estimator. All these tests are constructed in a univariate framework although some of them may be extended to multivariate continuous-time models.

Gallant and Tauchen (1996) propose a class of EMM tests that can be used to test multivariate continuous-time models. They propose a minimum χ^2 test for generic model misspecification and a class of diagnostic t-tests to gauge possible sources for model failure. Bhardwaj, Corradi, and Swanson (2008) consider a simulation-based test, which is an extension of the Andrews (1997) conditional

Kolmogorov test, for multivariate diffusion models. The limit distribution of their test is not nuisance parameter free, and asymptotic critical values must be obtained via a block bootstrap. Moreover, because these tests are by-products of the EMM and the simulated generalized method of moments (GMM) algorithms, respectively, they cannot be used when the model is estimated by other methods. This may limit the scope of these tests to otherwise very useful applications.

In a generalized cross-spectral non-Markov framework, Chen and Hong (2005) propose a new test for multivariate continuous-time models based on the conditional characteristic function (CCF), which often has a closed form or can be approximated accurately for many multifactor continuous-time models. As the Fourier transform of the transition density, the CCF contains the full information of the joint dynamics of underlying processes. This provides a basis for constructing an omnibus test for multifactor continuous-time models. Unlike Hong and Li (2005), Chen and Hong (2005) fully exploit the information in the joint transition density of underlying processes and hence can capture model misspecifications in their joint dynamics. Chen and Hong (2005) do not assume that the data generating process (DGP) is Markov. They take a generalized cross-spectral density approach, which employs many lags. For a Markov DGP (under both the null and alternative hypotheses), this test will not be most efficient, because it includes the past information of many lags, which is reductant under the Markov assumption. In this case, it is more efficient to focus on the first lag order only. This is pursued in the present paper.

There has been a long history of using the characteristic function in estimation and hypothesis testing. For example, Feuerverger and McDunnough (1981) discuss parameter estimation using the joint empirical characteristic function (ECF) for stationary Markov time series models. Epps and Pulley (1983) propose an omnibus test of normality via a weighted integral of the squared modulus of the difference between the characteristic functions of the observed sample and of the normal distribution. Su and White (2007) test conditional independence by comparing the unrestricted and restricted CCFs via a kernel regression. We note that all the preceding works deal with discrete-time models, but the characteristic function approach has attracted increasing attention in the continuous-time literature. For most continuous-time models, the transition density has no closed form, which makes estimation of and testing for continuous-time models rather challenging. However, for a general class of affine jump diffusion (AJD) models (e.g., Duffie, Pan, and Singleton, 2000) and time-changed Lévy processes (e.g., Chernov et al., 1999), the CCF has a closed form as an exponential affine function of state variables up to a system of ordinary differential equations. This fact has been exploited to develop new estimation methods for multifactor continuous-time models in the literature. Specifically, Chacko and Viceira (2003) suggest a spectral GMM estimator based on the average of the differences between the ECF and the model-implied characteristic function. Jiang and Knight (2002) derive the unconditional joint characteristic function of an AJD model and use it to develop some GMM and ECF estimation procedures. Singleton (2001) proposes both time-domain estimators based on the Fourier transform of the CCF and frequency-domain estimators directly based on the CCF. Carrasco, Chernov, Florens, and Ghysels (2007) propose GMM estimators with a continuum of moment conditions via the characteristic function. Besides its convenient closed form for many popular continuous-time models, the CCF can be differentiated to generate moments, which provides powerful and intuitive tools to check various specific aspects of a joint conditional distribution.

Motivated by these appealing features, we provide a CCF characterization for the adequacy of a continuous-time Markov model and use it to construct a specification test for continuous-time Markov models. The basic idea is that if a Markov model is correctly specified, prediction errors associated with the model-implied CCF should be a martingale difference sequence (MDS). This characterization has never been used in any goodness-of-fit test for continuous-time models, although it has been used in estimating them (e.g., Singleton, 2001). To ensure the power of our test, we use nonparametric regression to check whether these prediction errors are explainable by the current values of the underlying processes. Our approach has several attractive properties.

First, our omnibus test exploits the information in the joint transition density of state variables rather than only the information in the transition density of each component. Hence, it can capture various model misspecifications in the joint dynamics of state variables.[4] In particular, it can detect misspecifications in the joint transition density even if the transition density of each component is correctly specified.

Second, our test is applicable to a wide variety of continuous-time Markov models, such as diffusions, jump diffusions, and continuous-time Markov chains. Because we use the CCF, our test is most convenient when the model has a closed-form CCF, as is the case for AJD models (e.g., Duffie et al., 2000) and time-changed Lévy processes (e.g., Chernov et al., 1999). However, our test is also applicable to continuous-time Markov models with no closed-form CCF. In this case, we can use inverse Fourier transforms or simulation techniques to calculate the CCF. Moreover, our test is applicable to partially observed multifactor continuous-time Markov models. An example is the stochastic volatility (SV) models.

Third, we do not require any particular parameter estimation method. Any $\sqrt{T}$-consistent estimators may be used. This makes our test easily implementable in light of the notorious difficulty of obtaining asymptotically efficient estimators for multifactor continuous-time models. The inputs needed to calculate the test statistics are the observed data and the model-implied CCF or its approximation. Because we impose our conditions on the CCF of a discretely observed sample of a continuous-time Markov model, our test is also readily applicable to discrete-time Markov distribution models.

Fourth, in addition to the omnibus test, we also propose a class of diagnostic tests by differentiating the CCF. These derivative tests provide useful information on how well a continuous-time Markov model captures various specific aspects of the dynamics. In particular, they can reveal information on neglected dynamics in conditional means, conditional variances, and conditional correlations, respectively.

In Section 2, we introduce the hypotheses of interest and provide a characterization for correct specification of a continuous-time Markov model. In Section 3, we propose an omnibus goodness-of-fit test using smoothed regression, and in Section 4 we derive the asymptotic null distribution of our omnibus test and discuss its asymptotic power property. We then construct a class of diagnostic procedures that focus on various specific aspects of the joint dynamics of a multifactor continuous-time model in Section 5. In Section 6, we consider the tests for multifactor continuous-time models with partially unobservable components. In Section 7, we apply our tests to both univariate and bivariate continuous-time models in a simulation study. A conclusion follows in Section 8. All mathematical proofs are collected in an Appendix. Throughout, we will use C to denote a generic bounded constant, $\|\cdot\|$ for the euclidean norm, and A^* for the complex conjugate of A.

2. HYPOTHESES OF INTEREST

Given a complete probability space $(\Omega, \mathcal{F}, P)$ and an information filtration $\mathcal{F}_t$, we assume that a $d \times 1$ state vector $\mathbf{X}_t$ is a continuous-time Markov process in some state space $\mathbf{D} \subset \mathbb{R}^d$, where $d \geq 1$ is an integer. In finance, the following class $\mathcal{M}$ of continuous-time models is often used to capture the dynamics of $\mathbf{X}_t$:

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t, \boldsymbol{\theta})\,dt + \boldsymbol{\sigma}(\mathbf{X}_t, \boldsymbol{\theta})\,d\mathbf{W}_t + dJ_t(\boldsymbol{\theta}), \qquad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{2.1}$$

where $\mathbf{W}_t$ is a $d \times 1$ standard Brownian motion in $\mathbb{R}^d$, $\boldsymbol{\Theta}$ is a finite-dimensional parameter space, $\boldsymbol{\mu} : \mathbf{D} \times \boldsymbol{\Theta} \to \mathbb{R}^d$ is a drift function (i.e., instantaneous conditional mean), $\boldsymbol{\sigma} : \mathbf{D} \times \boldsymbol{\Theta} \to \mathbb{R}^{d\times d}$ is a diffusion function (i.e., instantaneous conditional standard deviation), and J_t is a pure jump process whose jump size follows a probability distribution $\nu : \mathbf{D} \times \boldsymbol{\Theta} \to \mathbb{R}^+$ and whose jump times arrive with intensity $\lambda : \mathbf{D} \times \boldsymbol{\Theta} \to \mathbb{R}^+$.[5]

The preceding setup is a general multifactor framework that nests most existing continuous-time Markov models in finance. For example, suppose the drift $\boldsymbol{\mu}(\cdot,\cdot)$, the instantaneous covariance matrix $\boldsymbol{\sigma}(\cdot,\cdot)\boldsymbol{\sigma}(\cdot,\cdot)'$, and the jump intensity $\lambda(\cdot,\cdot)$ are all affine functions of the state vector $\mathbf{X}_t$, namely,

$$\begin{cases} \boldsymbol{\mu}(\mathbf{X}_t, \boldsymbol{\theta}) = \mathbf{K}_0 + \mathbf{K}_1 \mathbf{X}_t, \\ [\boldsymbol{\sigma}(\mathbf{X}_t, \boldsymbol{\theta})\boldsymbol{\sigma}(\mathbf{X}_t, \boldsymbol{\theta})']_{jl} = [\mathbf{H}_0]_{jl} + [\mathbf{H}_1]_{jl}\mathbf{X}_t, \qquad 1 \leq j, l \leq d, \\ \lambda(\mathbf{X}_t, \boldsymbol{\theta}) = L_0 + \mathbf{L}_1 \mathbf{X}_t, \end{cases} \tag{2.2}$$

where $\mathbf{K}_0 \in \mathbb{R}^d$, $\mathbf{K}_1 \in \mathbb{R}^{d\times d}$, $\mathbf{H}_0 \in \mathbb{R}^{d\times d}$, $\mathbf{H}_1 \in \mathbb{R}^{d\times d\times d}$, $L_0 \in \mathbb{R}$, and $\mathbf{L}_1 \in \mathbb{R}^d$ are unknown parameters. Then we obtain the class of popular AJD models of Duffie et al. (2000).

It is well known that for a continuous-time Markov model described in (2.1), the specification of the drift $\boldsymbol{\mu}(\mathbf{X}_t, \boldsymbol{\theta})$, the diffusion $\boldsymbol{\sigma}(\mathbf{X}_t, \boldsymbol{\theta})$, and the jump process $J_t(\boldsymbol{\theta})$ together completely determines the joint transition density of the state vector $\mathbf{X}_t$. We use $p(\mathbf{x}, t|\mathbf{X}_s, \boldsymbol{\theta})$ to denote the model-implied transition density function of $\mathbf{X}_t = \mathbf{x}$ given $\mathbf{X}_s$, where $s < t$. Suppose $\mathbf{X}_t$ has an unknown true transition density $p_0(\mathbf{x}, t|\mathbf{X}_s)$. Then the continuous-time Markov model is correctly specified for the full dynamics of $\mathbf{X}_t$ if there exists some unknown parameter value $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ such that

$$\mathbb{H}_0 : p(\mathbf{x},t|\mathbf{X}_s, \boldsymbol{\theta}_0) = p_0(\mathbf{x}, t|\mathbf{X}_s) \quad \text{almost surely (a.s.)} \quad \text{and for all } t, s, s < t. \tag{2.3}$$

Alternatively, if for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we have

$$\begin{aligned} \mathbb{H}_A : {} & p(\mathbf{x}, t|\mathbf{X}_s, \boldsymbol{\theta}) \\ & \neq p_0(\mathbf{x}, t|\mathbf{X}_s) \quad \text{for some } t > s \text{ with positive probability measure,} \end{aligned} \tag{2.4}$$

then the continuous-time model is misspecified for the full dynamics of $\mathbf{X}_t$. We maintain the Markov assumption for $\mathbf{X}_t$ under both $\mathbb{H}_0$ and $\mathbb{H}_A$.

The transition density–based characterization can be used to test correct specification of the continuous-time model. When $\mathbf{X}_t$ is univariate, Hong and Li (2005) propose a test for a continuous-time model by checking whether the PIT

$$Z_t(\boldsymbol{\theta}_0) \equiv \int_{-\infty}^{\mathbf{X}_t} p(\mathbf{x}, t|\mathbf{X}_{t-\Delta}, \boldsymbol{\theta}_0)\, dx \sim i.i.d. U[0, 1] \quad \text{under } \mathbb{H}_0, \tag{2.5}$$

where Δ is the sampling interval for a discretely observed sample. The most appealing merit of this test is its robustness to persistent dependence in $\{\mathbf{X}_t\}$. However, there are some limitations to this approach. For example, for most continuous-time diffusion models (except some simple diffusion models such as the Vasicek, 1977, model), the transition densities have no closed form. Most importantly, the PIT cannot be applied to the multifactor joint transition density $p(\mathbf{x}, t|\mathbf{X}_{t-\Delta}, \boldsymbol{\theta})$, because when $d > 1$,

$$Z_t(\boldsymbol{\theta}_0) = \int_{-\infty}^{X_{1,t}} \cdots \int_{-\infty}^{X_{d,t}} p(\mathbf{x}, t|\mathbf{X}_{t-\Delta}, \boldsymbol{\theta}_0)\, d\mathbf{x} \tag{2.6}$$

is no longer i.i.d. $U[0,1]$ even if $\mathbb{H}_0$ holds, where $\mathbf{X}_t = (X_{1,t}, \ldots, X_{d,t})'$. Hong and Li (2005) suggest using the PIT for each state variable with a suitable partitioning. This is valid, but it does not make full use of the information contained in the joint distribution of $\mathbf{X}_t$. In particular, it may miss misspecifications in the joint dynamics of $\mathbf{X}_t$. For example, consider the DGP

$$d\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix} \begin{pmatrix} \theta_1 - X_{1,t} \\ \theta_2 - X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} d\begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix},$$

where $\left\{W_{1,t}, W_{2,t}\right\}$ are independent standard Brownian motions. Suppose we fit the data using the model

$$d\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{22} \end{pmatrix} \begin{pmatrix} \theta_1 - X_{1,t} \\ \theta_2 - X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} d\begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}.$$

This model is misspecified because it ignores correlations in drift. Now, following Hong and Li (2005), we calculate the generalized residuals $\{Z_t\} = \{Z_{1,t}, Z_{2,t}, Z_{1,t-\Delta}, Z_{2,t-\Delta}, \ldots\}$, where $Z_{1,t}$ and $Z_{2,t}$ are the PITs of $X_{1,t}$ and $X_{2,t}$ with respect to the conditional density models $p(X_{1,t}, t|\mathbf{X}_{t-\Delta}, X_{2,t}, \boldsymbol{\theta})$ and $p(X_{2,t}, t|\mathbf{X}_{t-\Delta}, \boldsymbol{\theta})$, respectively, and $\boldsymbol{\theta} = (\kappa_{11}, \kappa_{22}, \theta_1, \theta_2, \sigma_{11}, \sigma_{22})'$. Then the Hong and Li (2005) test has no power because each of these PITs is an i.i.d. $U[0,1]$ sequence.

As the Fourier transform of the transition density, the CCF can capture the full dynamics of $\mathbf{X}_t$. Let $\varphi(\mathbf{u}, t, |\mathbf{X}_s, \boldsymbol{\theta})$ be the model-implied CCF of $\mathbf{X}_t$, conditional on $\mathbf{X}_s$ at time $s < t$. That is,

$$\varphi(\mathbf{u}, t|\mathbf{X}_s, \boldsymbol{\theta}) \equiv \mathrm{E}_{\boldsymbol{\theta}}\left[\exp\left(i\mathbf{u}'\mathbf{X}_t\right)|\mathbf{X}_s\right] = \int_{\mathbb{R}^d} \exp\left(i\mathbf{u}'\mathbf{x}\right) p(\mathbf{x}, t|\mathbf{X}_s, \boldsymbol{\theta})\, d\mathbf{x},$$

$$\mathbf{u} \in \mathbb{R}^d, \qquad i = \sqrt{-1}, \tag{2.7}$$

where $\mathrm{E}_{\boldsymbol{\theta}}(\cdot|\mathbf{X}_s)$ denotes the expectation under the model-implied transition density $p(\mathbf{x}, t|\mathbf{X}_s, \boldsymbol{\theta})$.

Given the equivalence between the transition density and the CCF, the hypotheses of interest $\mathbb{H}_0$ in (2.3) versus $\mathbb{H}_A$ in (2.4) can be written as follows:

$$\mathbb{H}_0 : \mathrm{E}\left[\exp\left(i\mathbf{u}'\mathbf{X}_t\right)|\mathbf{X}_s\right]$$

$$= \varphi(\mathbf{u}, t|\mathbf{X}_s, \boldsymbol{\theta}_0) \text{ a.s.} \quad \text{for all } \mathbf{u} \in \mathbb{R}^d \quad \text{and for some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \tag{2.8}$$

versus

$$\mathbb{H}_A : \mathrm{E}\left[\exp\left(i\mathbf{u}'\mathbf{X}_t\right)|\mathbf{X}_s\right]$$

$$\neq \varphi(\mathbf{u}, t|\mathbf{X}_s, \boldsymbol{\theta}) \quad \text{with positive probability measure for all } \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{2.9}$$

Suppose we have a discretely observed sample $\{\mathbf{X}_t\}_{t=\Delta}^{T\Delta}$ of size T, where Δ is a fixed sampling interval. For simplicity we set $\Delta = 1$ in most places in the paper. Define a complex-valued process

$$Z_t(\mathbf{u}, \boldsymbol{\theta}) \equiv \exp\left(i\mathbf{u}'\mathbf{X}_t\right) - \varphi(\mathbf{u}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}), \qquad \mathbf{u} \in \mathbb{R}^d \quad \text{and} \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{2.10}$$

Then $\mathbb{H}_0$ is equivalent to the following MDS characterization:

$$\mathrm{E}\left[Z_t(\mathbf{u}, \boldsymbol{\theta}_0)|\mathbf{X}_{t-1}\right] = 0 \text{ a.s.} \quad \text{for all } \mathbf{u} \in \mathbb{R}^d \quad \text{and some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}. \tag{2.11}$$

It is important to emphasize that (2.11) is not a simple MDS characterization. It is a MDS process $Z_t(\cdot, \boldsymbol{\theta}_0)$ indexed by parameter $\mathbf{u} \in \mathbb{R}^d$, and we need to check all possible values for $\mathbf{u}$ in $\mathbb{R}^d$. This is challenging, but it offers the omnibus property of the resulting test. Moreover, by taking derivatives with respect to $\mathbf{u}$ at the origin, we can direct the test toward certain specific aspects of the joint dynamics of $\mathbf{X}_t$.

To compute $Z_t(\mathbf{u}, \boldsymbol{\theta})$, we need to know the CCF. In principle, we can always recover the CCF by simulation when it has no closed form. For a given $\boldsymbol{\theta}$ and conditional on $\mathbf{X}_{t-1}$, we can generate a large sequence $\{\tilde{\mathbf{X}}_{t|t-1}^{\boldsymbol{\theta},j}, j = 1, \ldots, J\}$ via, e.g., the Euler or generalized Milstein scheme (e.g., Kloeden, Platen, and Schurz, 1994) and then estimate the CCF by $\hat{\varphi}(\mathbf{u}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}) = 1/J \sum_{j=1}^{J} \exp\left(i\mathbf{u}\tilde{\mathbf{X}}_{t|t-1}^{\boldsymbol{\theta},j}\right)$. It can be shown that for each t, $\hat{\varphi}(\mathbf{u}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}) - \varphi(\mathbf{u}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}) \to^{P} 0$ if $J \to \infty$. Therefore, our CCF approach is generally applicable. Alternatively, we can accurately approximate the model transition density by using, e.g., the Hermite expansion method of Aït-Sahalia (2002), the simulation methods of Brandt and Santa-Clara (2002) and Pedersen (1995), or the closed form approximation method of Duffie, Pedersen, and Singleton (2003) and then calculate the Fourier transform of the estimated transition density. Nevertheless, our test is most useful when the CCF has a closed form, as is illustrated by the examples that follow.

AJD models are a class of important continuous-time models with a closed-form CCF (Dai and Singleton, 2000; Duffie and Kan, 1996; Duffie et al., 2000). It has been shown (e.g., Duffie et al., 2000) that for AJD models, the CCF of $\mathbf{X}_t$ conditional on $\mathbf{X}_{t-1}$ is a closed-form exponential-affine function of $\mathbf{X}_{t-1}$:

$$\varphi(\mathbf{u}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}) = \exp\left[\alpha_{t-1}(\mathbf{u}) + \boldsymbol{\beta}_{t-1}(\mathbf{u})'\mathbf{X}_{t-1}\right], \tag{2.12}$$

where $\alpha_{t-1} : \mathbb{R}^d \to \mathbb{R}$ and $\boldsymbol{\beta}_{t-1} : \mathbb{R}^d \to \mathbb{R}^d$ satisfy the complex-valued Riccati equations

$$\begin{cases} \dot{\boldsymbol{\beta}}_t = \mathbf{K}_1'\boldsymbol{\beta}_t + \frac{1}{2}\boldsymbol{\beta}_t'\mathbf{H}_1\boldsymbol{\beta}_t + \mathbf{L}_1\left(g\left(\boldsymbol{\beta}_t\right) - 1\right), \\ \dot{\alpha}_t = \mathbf{K}_0'\boldsymbol{\beta}_t + \frac{1}{2}\boldsymbol{\beta}_t'\mathbf{H}_0\boldsymbol{\beta}_t + L_0\left(g\left(\boldsymbol{\beta}_t\right) - 1\right), \end{cases} \tag{2.13}$$

with boundary conditions $\boldsymbol{\beta}_T(\mathbf{u}) = i\mathbf{u}$ and $\alpha_T(\mathbf{u}) = 0$.

AJD models have been widely used in finance. For example, in the interest rate term structure literature, Dai and Singleton (2000), Duffie and Kan (1996), and Duffie et al. (2000) have developed a class of ATSMs. Assuming that the spot rate r_t is an affine function of the state vector $\mathbf{X}_t$ and that $\mathbf{X}_t$ follows affine diffusions under an equivalent martingale measure, Duffie and Kan (1996) show that the yield of the zero coupon bond

$$Y\left(\mathbf{X}_t, \tau\right) \equiv -\frac{1}{\tau}\log P\left(\mathbf{X}_t, \tau\right) = \frac{1}{\tau}\left[-A\left(\tau\right) + \mathbf{B}\left(\tau\right)'\mathbf{X}_t\right], \tag{2.14}$$

where τ is the remaining time to maturity and the functions A: $\mathbb{R}^+ \to \mathbb{R}$ and $\mathbf{B}$: $\mathbb{R}^+ \to \mathbb{R}^d$ either have a closed form or can be easily solved via numerical methods. Because $Y(\mathbf{X}_t, \tau)$ is a linear transformation of $\mathbf{X}_t$, its CCF also has the closed-form solution, namely,

$$\begin{aligned}\varphi_Y\left(u, t|\mathbf{X}_{t-1,}\boldsymbol{\theta}, \tau\right) &= \mathrm{E}_{\boldsymbol{\theta}}\{\exp[iuY(\mathbf{X}_t, \tau)]|Y(\mathbf{X}_{t-1}, \tau)\} \\ &= \exp\left\{-\frac{iu\mathbf{A}(\tau)}{\tau} + \alpha_{t-1}\left[\frac{u\mathbf{B}(\tau)}{\tau}\right] + \boldsymbol{\beta}_{t-1}\left[\frac{u\mathbf{B}(\tau)}{\tau}\right]'\mathbf{X}_{t-1}\right\},\end{aligned} \tag{2.15}$$

where α_{t-1} and $\boldsymbol{\beta}_{t-1}$ satisfy (2.13) and $\mathbf{X}_t = [\mathbf{B}(\tau)']^{-1}[\tau Y(\mathbf{X}_t, \tau) + A(\tau)]$. In particular, for a multifactor Vasicek model, the CCF of the yield of the zero coupon bond has an analytical expression.

Assuming that the spot rate is a quadratic function of the normally distributed state vector, Ahn, Dittmar, and Gallant (2002) derive the yield of the zero coupon bond as a quadratic function of the state vector $\mathbf{X}_t$:

$$Y(\mathbf{X}_t, \tau) \equiv -\frac{1}{\tau}\log P(\mathbf{X}_t, \tau) = \frac{1}{\tau}\left[-A(\tau) - \mathbf{B}(\tau)'\mathbf{X}_t - \mathbf{X}_t'\mathbf{M}(\tau)\mathbf{X}_t\right], \tag{2.16}$$

where functions $A : \mathbb{R}^+ \to \mathbb{R}$, $\mathbf{B} : \mathbb{R}^+ \to \mathbb{R}^d$, and $\mathbf{M} : \mathbb{R}^+ \to \mathbb{R}^{d\times d}$ either have a closed form or can be easily solved via numerical methods. This class of models is called the quadratic term structure models (QTSMs), for which the CCF of $Y(\mathbf{X}_t, \tau)$ is

$$\begin{aligned}\varphi_Y(u, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}, \tau) &= \exp\left\{iu\left[-\frac{A}{\tau} + \frac{\mathbf{B}'\left(\mathbf{M}^{-1}\right)'\mathbf{B}}{4\tau}\right] + \sum_{j=1}^{d}\frac{iu\lambda_j\omega_j^2}{1-2iu\lambda_j}\right\} \\ &\quad\times\prod_{j=1}^{d}\left(1-2iu\lambda_j\right)^{-1/2},\end{aligned} \tag{2.17}$$

where λ_j and ω_j $(j = 1, 2, \ldots, d)$ are some constants defined in Ahn et al. (2002).

3. NONPARAMETRIC REGRESSION-BASED CCF TESTING

We now propose a test for the adequacy of a multivariate continuous-time Markov model using nonparametric regression. Recall that the CCF-based generalized residual $Z_t(\mathbf{u}, \boldsymbol{\theta})$ in (2.10) has the MDS property:

$$\begin{aligned}&m\left(\mathbf{u}, \mathbf{X}_{t-1}, \boldsymbol{\theta}_0\right) \\ &\quad\equiv \mathrm{E}\left[Z_t(\mathbf{u}, \boldsymbol{\theta}_0)|\mathbf{X}_{t-1}\right] = 0 \text{ a.s.} \quad \text{for all } \mathbf{u} \in \mathbb{R}^d \quad \text{and some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}.\end{aligned}$$

To gain insight into the MDS characterization for $Z_t(\mathbf{u},\boldsymbol{\theta}_0)$, we take a Taylor series expansion of $m(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta})$ with respect to $\mathbf{u}$ around zero. This yields

$$m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}) = \sum_{|\boldsymbol{\nu}|=0}^{\infty} \frac{m^{(\nu)}(\mathbf{0},\mathbf{x},\boldsymbol{\theta})}{\Pi_{c=1}^{d}\nu_c!} u_1^{\nu_1}\ldots u_d^{\nu_d}, \quad \text{where } \mathbf{u} = (u_1,\ldots,u_d)',$$

$$\begin{aligned} m^{(\boldsymbol{\nu})}(\mathbf{0},\mathbf{x},\boldsymbol{\theta}) &\equiv \frac{\partial^{\nu_1}}{\partial u_1^{\nu_1}}\cdots\frac{\partial^{\nu_d}}{\partial u_d^{\nu_d}} m(\mathbf{u},\mathbf{x},\boldsymbol{\theta})\bigg|_{\mathbf{u}=\mathbf{0}} \\ &= \mathrm{E}\left[\prod_{c=1}^{d}(iX_{c,t})^{\nu_c}|\mathbf{X}_{t-1}\right] - \mathrm{E}_{\boldsymbol{\theta}}\left[\prod_{c=1}^{d}(iX_{c,t})^{\nu_c}|\mathbf{X}_{t-1}\right]. \end{aligned}$$

Here, as before, $\mathrm{E}_{\boldsymbol{\theta}}(\cdot|\mathbf{X}_{t-1})$ is the expectation under the model-implied transition density $p(\mathbf{x},t|\mathbf{X}_{t-1},\boldsymbol{\theta}), \boldsymbol{\nu} = (\nu_1,\nu_2,\ldots,\nu_d)'$, and $|\boldsymbol{\nu}| = \sum_{c=1}^{d}\nu_c$. Thus, checking the MDS condition for $Z_t(\mathbf{u},\boldsymbol{\theta}_0)$ is equivalent to checking whether the dynamics of various conditional moments and cross-moments of $\mathbf{X}_t$ have been adequately captured by the null continuous-time model. The MDS characterization thus provides a novel approach to constructing an omnibus test that does not have to use various conditional moments and cross-moments of $\mathbf{X}_t$. This is particularly appealing when higher order moments of financial time series may not exist.

Given a discretely observed sample $\{\mathbf{X}_t\}_{t=1}^{T}$, we can estimate the complex-valued regression function $m(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)$ nonparametrically and check whether $m(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)$ is identically zero for all $\mathbf{u}\in\mathbb{R}^d$ and some $\boldsymbol{\theta}_0\in\boldsymbol{\Theta}$. Nonparametric estimation of $m(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)$ is suitable here because $m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}_0)$ is potentially highly nonlinear under $\mathbb{H}_A$. Various nonparametric regression methods could be used. We use local linear regression here. Local linear regression is introduced by Stone (1977) and studied by Cleveland (1979) and Fan (1993), among many others. It has significant advantages over the conventional Nadaraya–Watson estimator. It reduces the bias and adapts automatically to the boundary of design points (see Fan and Gijbels, 1996). Fan (1993) shows that within the class of linear estimators that includes kernel and spline estimators, the local linear estimators achieve the best possible rates of convergence.

We consider the following local least squares problem:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{d+1}} \sum_{t=2}^{T} \left|Z_t(\mathbf{u},\boldsymbol{\theta}_0) - \beta_0 - \boldsymbol{\beta}_1'(\mathbf{X}_t-\mathbf{x})\right|^2 K_h(\mathbf{x}-\mathbf{X}_t),$$
$$\mathbf{x}\in\mathbb{R}^d, \qquad \mathbf{u}\in\mathbb{R}^d, \qquad (3.1)$$

where $\boldsymbol{\beta} = \left(\beta_0,\boldsymbol{\beta}_1'\right)'$ is a $(d+1)\times 1$ parameter vector, $K_h(\mathbf{x}) = h^{-d}K(\mathbf{x}/h)$, $K:\mathbb{R}^d\to\mathbb{R}$ is a kernel, and h is a bandwidth. An example of $K(\cdot)$ is a symmetric probability density. We obtain the following solution:

$$\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\mathbf{x},\mathbf{u}) = \begin{bmatrix}\hat{\beta}_0(\mathbf{x},\mathbf{u})\\ \hat{\boldsymbol{\beta}}_1(\mathbf{x},\mathbf{u})\end{bmatrix} = [\mathbf{X}'\mathbf{W}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}\mathbf{Z}, \qquad \mathbf{x}\in\mathbb{R}^d, \qquad (3.2)$$

where $\mathbf{X}$ is a $T \times (d+1)$ matrix with the tth row given by $[1, (\mathbf{X}_t - \mathbf{x})']$, $\mathbf{W} = \text{diag}[K_h(\mathbf{X}_1 - \mathbf{x}), \ldots, K_h(\mathbf{X}_T - \mathbf{x})]$, $\mathbf{Z} = [Z_1(\mathbf{u}, \boldsymbol{\theta}_0), \ldots, Z_T(\mathbf{u}, \boldsymbol{\theta}_0)]'$. Note that $\hat{\boldsymbol{\beta}}$ depends on the location $\mathbf{x}$ and parameter $\mathbf{u}$.

The function $m(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}_0)$ can be estimated by the local intercept estimator $\hat{\beta}_0(\mathbf{x}, \mathbf{u})$. Specifically,

$$\hat{m}\left(\mathbf{u}, \mathbf{x}, \hat{\boldsymbol{\theta}}\right) = \sum_{t=2}^{T} \hat{W}\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right) Z_t\left(\mathbf{u}, \hat{\boldsymbol{\theta}}\right), \tag{3.3}$$

where $\hat{W}(\cdot)$ is an effective kernel, defined as

$$\hat{W}(t) \equiv \mathbf{e}_1' \mathbf{S}_T^{-1} [1, th, \ldots, th]' K(t) / h^d, \tag{3.4}$$

$\mathbf{e}_1 = (1, 0, \ldots, 0)'$ is a $(d+1) \times 1$ unit vector, $\mathbf{S}_T = \mathbf{X}'\mathbf{W}\mathbf{X}$ is a $(d+1) \times (d+1)$ matrix. As established by Hjellvik, Yao, and Tjøstheim (1998), for any compact set $\mathbf{G} \subset \mathbb{R}^d$, one has $\mathbf{S}_T^{-1} = g(\mathbf{x})^{-1} \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{S}_0 \end{bmatrix}^{-1} + o_P(1)$ uniformly for $\mathbf{x} \in \mathbf{G}$, where $g(\mathbf{x})$ is the true stationary density of $\mathbf{X}_t$, $\mathbf{0}$ is a $d \times 1$ vector of zeros, and $\mathbf{S}_0$ is the $d \times d$ diagonal matrix whose diagonal element is $\int_{\mathbb{R}^d} \mathbf{u}\mathbf{u}' K(\mathbf{u})\, d\mathbf{u}$. It follows that the effective kernel

$$\hat{W}(t) = \frac{1}{Th^d g(\mathbf{x})} K(t) [1 + o_P(1)]. \tag{3.5}$$

Equation (3.5) shows that the local linear estimator behaves like a kernel regression estimator based on the kernel $K(\cdot)$ with a known design density. Under certain conditions, $\hat{m}(\mathbf{u}, \mathbf{x}, \hat{\boldsymbol{\theta}})$ is consistent for $m(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}_0)$. It converges to a zero function under $\mathbb{H}_0$ and a nonzero function under $\mathbb{H}_A$. We can measure the distance between $\hat{m}(\mathbf{u}, \mathbf{x}, \hat{\boldsymbol{\theta}})$ and a zero function by the quadratic form

$$L^2(\hat{m}) = \sum_{t=2}^{T} \int \left|\hat{m}(\mathbf{u}, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}})\right|^2 a(\mathbf{X}_{t-1})\, dW(\mathbf{u}), \tag{3.6}$$

where $a : \mathbb{R}^d \to \mathbb{R}^+$ is a weighting function and $W : \mathbb{R}^d \to \mathbb{R}^+$ is a nondecreasing weighting function that weighs sets symmetric about the origin equally.[6] The use of $a(\mathbf{X}_{t-1})$ is not uncommon in the literature; see, e.g., Aït-Sahalia, Bickel, and Stoker (2001), Aït-Sahalia et al. (2009), Hjellvik et al. (1998), and Su and White (2007). This is often used to remove extreme observations. As noted by Aït-Sahalia et al. (2001), by choosing an appropriate $a(\cdot)$, one can focus on a particular empirical question of interest and reduce the influences of unreliable estimates. On the other hand, to ensure omnibus power, we have to consider many points for $\mathbf{u}$. An example of $W(\cdot)$ is the $N(\mathbf{0}, \mathbf{I}_d)$ cumulative distribution function (cdf), where $\mathbf{I}_d$ is a $d \times d$ identity matrix. Note that $W(\cdot)$ need not be continuous. They can be nondecreasing step functions such as discrete multivariate cdfs. This is equivalent to using finitely many or countable grid points for $\mathbf{u}$.

The omnibus test statistic for $\mathbb{H}_0$ against $\mathbb{H}_A$ is an appropriately standardized version of (3.6):

$$\hat{M} = \left[h^{d/2} \sum_{t=2}^{T} \int \left| \hat{m}(\mathbf{u}, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}}) \right|^2 a\left(\mathbf{X}_{t-1}\right) dW(\mathbf{u}) - \hat{C} \right] \Big/ \sqrt{2\hat{D}}, \quad \text{where}$$

$$\hat{C} = h^{-d/2} \iint \left[1 - \left| \varphi\left(\mathbf{u}, t | \mathbf{x}, \hat{\boldsymbol{\theta}} \right) \right|^2 \right] a(\mathbf{x})\, d\mathbf{x}\, dW(\mathbf{u}) \int K^2(\boldsymbol{\tau})\, d\boldsymbol{\tau},$$

$$\hat{D} = \iiint \left| \varphi\left(\mathbf{u}+\mathbf{v}, t | \mathbf{x}, \hat{\boldsymbol{\theta}} \right) - \varphi\left(\mathbf{u}, t | \mathbf{x}, \hat{\boldsymbol{\theta}} \right) \varphi\left(\mathbf{v}, t | \mathbf{x}, \hat{\boldsymbol{\theta}} \right) \right|^2$$
$$\times a^2(\mathbf{x})\, d\mathbf{x}\, dW(\mathbf{u})\, dW(\mathbf{v}) \int \left[\int K(\boldsymbol{\tau}) K(\boldsymbol{\tau}+\boldsymbol{\eta})\, d\boldsymbol{\tau} \right]^2 d\boldsymbol{\eta}. \qquad \textbf{(3.7)}$$

The factors $\hat{C}$ and $\hat{D}$ are the approximate mean and variance of the quadratic form in (3.6).

In practice, $\hat{M}$ has to be calculated using numerical integration or approximated by simulation techniques. This may be computationally costly when the dimension d of $\mathbf{X}_t$ is large. Alternatively, one can only use a finite number of grid points for $\mathbf{u}$. For example, we can symmetrically generate finitely many numbers of $\mathbf{u}$ from an $N(\mathbf{0}, \mathbf{I}_d)$ distribution. This will reduce the computational cost but may lead to some power loss.

Both $\hat{C}$ and $\hat{D}$ are derived under $\mathbb{H}_0$ using an asymptotic argument. They may not approximate well the mean and variance of the quadratic form in (3.6). This may lead to poor size in finite samples, although not necessarily poor power. Alternatively, we also consider the following test statistic:

$$\hat{M}_{FS} = \left[h^{d/2} \sum_{t=2}^{T} \int \left| \hat{m}(\mathbf{u}, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}}) \right|^2 a\left(\mathbf{X}_{t-1}\right) dW(\mathbf{u}) - \hat{C}_{FS} \right] \Big/ \sqrt{2\hat{D}}, \quad \text{where}$$

$$\hat{C}_{FS} = h^{d/2} \sum_{s=2}^{T} \int \left| \mathbf{Z}_s\left(\mathbf{u}, \hat{\boldsymbol{\theta}} \right) \right|^2 dW(\mathbf{u}) \sum_{t=2}^{T} \hat{W}^2 \left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h} \right) a\left(\mathbf{X}_{t-1}\right) \qquad \textbf{(3.8)}$$

is a finite-sample version of $\hat{C}$. It is expected to give better approximation for the mean of $L^2(\hat{m})$ in finite samples. Similarly, we could also replace the scaling factor $\hat{D}$ by its finite-sample version

$$h^{d/2} \sum_{s=2}^{T} \sum_{r=2}^{s-1} \left\{ \sum_{t=2}^{T} \int \operatorname{Re}\left[\mathbf{Z}_s\left(\mathbf{u}, \hat{\boldsymbol{\theta}} \right) \mathbf{Z}_r^*\left(\mathbf{u}, \hat{\boldsymbol{\theta}} \right) \right] dW \right.$$
$$\left. \times\, (\mathbf{u}) \hat{W}\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h} \right) \hat{W}\left(\frac{\mathbf{X}_{r-1} - \mathbf{X}_{t-1}}{h} \right) a(\mathbf{X}_{t-1}) \right\}^2,$$

but its computational cost is rather substantial when the sample size T is large.

We emphasize that although the CCF and the transition density are Fourier transforms of each other, our nonparametric regression-based CCF approach has an advantage over the nonparametric transition density–based approach that compares a nonparametric transition density with the model-implied transition density $p(\mathbf{x},t|\mathbf{X}_s,\hat{\boldsymbol{\theta}})$ via a quadratic form (e.g., Aït-Sahalia et al., 2009). This follows because our nonparametric regression estimator in (3.3) is only d-dimensional but the nonparametric transition density estimator is $2d$-dimensional. We expect that such dimension reduction will give better size and power in finite samples.

4. ASYMPTOTIC THEORY

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

Assumption A.1. Let $(\Omega,\mathcal{F},P)$ be a complete probability space.

(a) The stochastic time series vector process $\mathbf{X}_t \equiv \mathbf{X}_t(\omega)$, where $\omega \in \Omega$ and $t \in [0,T] \subset \mathbb{R}^+$, is a $d \times 1$ strictly stationary continuous-time Markov process with the marginal density $g(\mathbf{x})$, which is positive and continuous for all $\mathbf{x} \in \mathbf{G}$, where $\mathbf{G}$ is a compact set of $\mathbb{R}^d$. Also, the joint density of $(\mathbf{X}_1,\mathbf{X}_l)$ is continuous and bounded by some constant independent of $l > 1$.

(b) A discrete sample $\{\mathbf{X}_t\}_{t=\Delta}^{T\Delta}$,where $\Delta \equiv 1$ is the sampling interval, is observed at equally spaced discrete times, and $\{\mathbf{X}_t\}_{t=\Delta}^{T\Delta}$ is a β-mixing process with mixing coefficients satisfying $\sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} < C$ for some $0 < \delta < 1$.

Assumption A.2. Let $\varphi(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta})$ be the CCF of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$ of a continuous-time Markov parametric model $\mathcal{M} = \mathcal{M}(\boldsymbol{\theta})$ indexed by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(a) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, each $\mathbf{u} \in \mathbb{R}^d$, and each t, $\varphi(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta})$ is measurable with respect to $\mathbf{X}_{t-1}$.

(b) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, each $\mathbf{u} \in \mathbb{R}^d$, and each t, $\varphi(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with probability one; and

(c) $\sup_{\mathbf{u}\in\mathbb{R}^d} E\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\partial}{\partial\boldsymbol{\theta}}\varphi(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta})\right\|^2 \le C$ and $\sup_{\mathbf{u}\in\mathbb{R}^d} E\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\varphi(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta})\right\| \le C$.

Assumption A.3. $\hat{\boldsymbol{\theta}}$ is a parameter estimator such that $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = O_P(1)$.

Assumption A.4. The function $K : \mathbb{R}^d \to \mathbb{R}^+$ is a product kernel of some univariate kernel k, i.e., $K(\mathbf{u}) = \prod_{j=1}^d k(u_j)$, where $k : \mathbb{R} \to \mathbb{R}^+$ satisfies the Lipschitz condition and is a symmetric, bounded, and twice continuously differentiable function with $\int_{-\infty}^{\infty} k(u)\,du = 1, \int_{-\infty}^{\infty} uk(u)\,du = 0,$ and $\int_{-\infty}^{\infty} u^2 k(u)\,du < \infty$.

Assumption A.5.

(a) $W : \mathbb{R}^d \to \mathbb{R}^+$ is a nondecreasing right-continuous weighting function that weighs sets symmetric about the origin equally, with $\int_{\mathbb{R}^d} \|\mathbf{u}\|^4 dW(\mathbf{u}) < \infty$;

(b) $a : \mathbf{G} \to \mathbb{R}^+$ is a bounded weighting function that is continuous over $\mathbf{G}$, where $\mathbf{G} \in \mathbb{R}^d$ is the compact support given in Assumption A.1.

Assumption A.1 imposes regularity conditions on the DGP. Both univariate and multivariate continuous-time processes are covered. Following Aït-Sahalia (1996a, 1996b), Gallant and Long (1997), and Gallant and Tauchen (1996), we impose regularity conditions on a discretely observed random sample. There are two kinds of asymptotic results in the literature. The first is to let the sampling interval $\Delta \to 0$. This implies that the number of observations per unit of time tends to infinity. The second is to let the time horizon $T \to \infty$. As argued by Aït-Sahalia (1996b), the first approach hardly matches the way in which new data are added to the sample. Moreover, even if such ultra high–frequency data are available, market microstructural problems are likely to complicate the analysis considerably. Hence, like Aït-Sahalia (1996a) and Singleton (2001), we fix the sampling interval Δ and derive the asymptotic properties of our test for an expanding sampling period. Unlike Aït-Sahalia (1996a, 1996b), however, we avoid imposing additional assumptions on the stochastic differential equation (SDE), because we consider a more general framework. We allow but do not assume $\mathbf{X}_t$ to be a diffusion process.

We assume that the DGP is Markov under both $\mathbb{H}_0$ and $\mathbb{H}_A$ and focus on testing functional form misspecification. Given the fact that the Markov is a maintained condition for almost all continuous-time models (e.g., diffusion, jump diffusion, and Lévy processes), if these continuous-time models are correctly specified, the Markov assumption of the DGP is satisfied under $\mathbb{H}_0$. Hence our approach is applicable to these models.

The β-mixing condition restricts the degree of temporal dependence in $\{\mathbf{X}_t\}$. We say that $\mathbf{X}_t$ is β-mixing if $\beta(j) = \sup_{s \geq 1} \mathrm{E}\left[\sup_{A \in \mathcal{F}_{s+j}^{\infty}} \left|P\left(A|\mathcal{F}_1^s\right) - P(A)\right|\right] \to 0$, as $j \to \infty$, where $\mathcal{F}_j^s$ is the σ-field generated by $\{\mathbf{X}_\tau : \tau = j, \ldots, s\}$, $j \leq s$. Aït-Sahalia et al. (2009), Hjellvik et al. (1998), and Su and White (2007) also impose β -mixing conditions in related contexts. Our mixing condition is weaker than that of Aït-Sahalia et al. (2009), who assume β-mixing with an exponential decay rate. Suggested by Hansen and Scheinkman (1995) and Aït-Sahalia (1996a), one set of sufficient conditions for the β-mixing when $d = 1$ is (a) $\lim_{x \to l \text{or} x \to u} \sigma(x, \boldsymbol{\theta}) \pi(x, \boldsymbol{\theta}) = 0$; and (b) $\lim_{x \to l \text{or} x \to u} |\sigma(x, \boldsymbol{\theta}) / \{2\mu(x, \boldsymbol{\theta}) - \sigma(x, \boldsymbol{\theta})[\partial\sigma(x, \boldsymbol{\theta})/\partial x]\}| < \infty$, where l and u are left and right boundaries of X_t with possibly $l = -\infty$ and/or $u = +\infty$ and where $\pi(x, \boldsymbol{\theta})$ is the model-implied marginal density.

Assumption A.2 provides conditions on multifactor continuous-time Markov models. We impose these conditions directly on the model-implied CCF, which covers other continuous-time processes not characterized by a SDE. As the CCF is the Fourier transform of the transition density, we can easily translate the conditions on the model-implied CCF into the conditions on the model-implied transition density $p(\mathbf{x}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta})$. In particular, Assumption A.2 holds if (a) for each t, each $\mathbf{x} \in \mathbf{G}$, and each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $p(\mathbf{x}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta})$ is measurable with respect to $\mathbf{X}_{t-1}$; (b) $p(\mathbf{x}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with probability one; (c) $\sup_{\mathbf{x}\in\mathbf{G}} \mathrm{E} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial\boldsymbol{\theta}} \ln p(\mathbf{x}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}) \right\|^2 \leq C$ and $\sup_{\mathbf{x}\in\mathbf{G}} \mathrm{E} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \ln p(\mathbf{x}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}) \right\| \leq C$. An advantage of imposing conditions on the model-implied CCF or transition density is that the asymptotic theory of our tests is readily applicable to test the validity of a discrete-time conditional distribution model.

Assumption A.3 requires a $\sqrt{T}$-consistent estimator $\hat{\boldsymbol{\theta}}$ under $\mathbb{H}_0$. We allow using both asymptotically optimal and suboptimal estimators, such as the Aït-Sahalia (2002) approximated MLE, the Chib et al. (2004) MCMC method, the Gallant and Tauchen (1996) EMM, the Singleton (2001) MLE-CCF and GMM-CCF, and the quasi-MLE. We do not require any asymptotically most efficient estimator or a specified estimator. This is attractive given the notorious difficulty of asymptotically efficient estimation of multifactor continuous-time models and may be viewed as an advantage over some existing tests that require a specific estimation method.

Assumption A.5 imposes some mild conditions on the weighting functions $W(\mathbf{u})$ and $a(\mathbf{x})$, respectively. Any cdf with a finite fourth moment satisfies the condition for $W(\mathbf{u})$. The function $W(\mathbf{u})$ need not be continuous. This provides a convenient way to implement our test, because we can avoid high-dimensional numerical integrations by using finitely many or countable grid points for $\mathbf{u}$. For simplicity of the proof, we assume that the weighting function $a(\mathbf{x})$ has a compact support on $\mathbb{R}^d$.

We now state the asymptotic distribution of $\hat{M}$ under $\mathbb{H}_0$.

THEOREM 1. *Suppose Assumptions A.1–A.5 hold and* $h = cT^{-\lambda}$ *for* $0 < \lambda < \frac{1}{2d}$ *and* $0 < c < \infty$. *Then* $\hat{M} \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0$ *as* $T \to \infty$.

As an important feature of $\hat{M}$, the use of the estimated generalized residuals $\{Z_t(\mathbf{u},\hat{\boldsymbol{\theta}})\}$ in place of the true residuals $\{Z_t(\mathbf{u},\boldsymbol{\theta}_0)\}$ has no impact on the limit distribution of $\hat{M}$. One can proceed as if the true parameter value $\boldsymbol{\theta}_0$ were known and equal to $\hat{\boldsymbol{\theta}}$. Intuitively, the parametric estimator $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$ faster than the nonparametric estimator $\hat{m}(\mathbf{u},\mathbf{x},\boldsymbol{\theta}_0)$ to $m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}_0)$. Consequently, the limit distribution of $\hat{M}$ is solely determined by $\hat{m}(\mathbf{u},\mathbf{x},\boldsymbol{\theta}_0)$, and replacing $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}}$ has no impact asymptotically.[7] This delivers a convenient procedure, because any $\sqrt{T}$-consistent estimator can be used.

Theorem 1 allows a wide range of admissible rates for nonstochastic bandwidth h. In practice, one might like to choose h via some data-driven method, which can let data determine an appropriate lag order. For example, an automatic method such as the Fan and Gijbels (1996) plug-in method may be used. A possible choice of $\hat{h}$ is

$$\hat{h} = \arg\min_h IMSE(h) \equiv \int\int \left[\hat{B}\left(\mathbf{u},\mathbf{x},\hat{\boldsymbol{\theta}}\right)^2 + \hat{V}\left(\mathbf{u},\mathbf{x},\hat{\boldsymbol{\theta}}\right)\right] a(\mathbf{x})\, d\mathbf{x}\, dW(\mathbf{u}),$$

where $\hat{B}(\mathbf{u},\mathbf{x},\hat{\boldsymbol{\theta}})$ and $\hat{V}(\mathbf{u},\mathbf{x},\hat{\boldsymbol{\theta}})$ are the estimated bias and variance of $\hat{m}(\mathbf{u},\mathbf{x},\hat{\boldsymbol{\theta}})$, respectively.

Another example is the cross-validation method to choose $\hat{h}$, namely,

$$\hat{h}_{CV} = \arg\min_h CV(h) \equiv \sum_{t=2}^{T} \int \left| Z_t\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right) - \hat{m}^-\left(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}}\right)\right|^2 a\left(\mathbf{X}_{t-1}\right) dW(\mathbf{u}),$$

where $\hat{m}^-\left(\mathbf{u},\mathbf{X}_t,\hat{\boldsymbol{\theta}}\right) = \sum_{s=1,s\neq t}^{T} \hat{W}\left((\mathbf{X}_s - \mathbf{X}_t)/h\right) Z_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)$ is a "leave-one-out" estimator.

To justify the use of a data-driven bandwidth $\hat{h}$, we impose the following condition on the kernel $k(\cdot)$.

Assumption A.6. The univariate kernel $k(\cdot)$ is three times differentiable with $\int \left|x^s k^{(s)}(x)\right| dx < \infty$ for all $s = 1,2,3,$where $k^{(s)}(\cdot)$ denotes the sth-order derivative of $k(\cdot)$.

Examples include the polynomial kernel class $c_p(1-u^2)^p$ for $p \geq 2$ (Müller, 1984) and the normal kernel.

THEOREM 2. *Suppose Assumptions A.1–A.6 hold, $\hat{h}$ is a random bandwidth such that $(\hat{h}-h)/h = o_P\left(h^{d/2}\right)$, where nonstochastic bandwidth $h = cT^{-\lambda}$ for $0 < \lambda < \frac{1}{2d}$ and $0 < c < \infty$. Then $\hat{M}_{\hat{h}} \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$ as $T \to \infty$, where $\hat{M}_{\hat{h}}$ is computed in the same way as $\hat{M}$, with $\hat{h}$ replacing h.*

The use of $\hat{h}$ has an asymptotically negligible impact on the limit distribution of $\hat{M}_{\hat{h}}$ provided that $\hat{h}/h \to 1$ in probability at a proper rate. The convergence rate condition on the random $\hat{h}$ to h is not restrictive. For example, suppose $h \propto T^{-1/(4+d)}$. Then we require $(\hat{h}-h)/h = o_P\left(T^{-d/2(d+4)}\right)$.

Next, we investigate the asymptotic power property of $\hat{M}$ under $\mathbb{H}_A$.

Assumption A.7. $\hat{\boldsymbol{\theta}}$ is a parameter estimator with $p\lim_{T\to\infty} \hat{\boldsymbol{\theta}} = \theta^* \in \boldsymbol{\Theta}$.

THEOREM 3. *Suppose Assumptions A.1–A.2, A.4, A.5, and A.7 hold and $h = cT^{-\lambda}$ for $0 < \lambda < \frac{2}{3d}$ and $0 < c < \infty$.*

(i) Then $T^{-1}h^{-d/2}\hat{M} \xrightarrow{p} (2D)^{-1/2} \iint \left|m\left(\mathbf{u},\mathbf{x},\boldsymbol{\theta}^*\right)\right|^2 a(\mathbf{x})g(\mathbf{x})\, d\mathbf{x}\, dW(\mathbf{u})$ as $T \to \infty$, where

$$D = \iiint \left|\varphi\left(\mathbf{u}+\mathbf{v},t|\mathbf{x},\boldsymbol{\theta}^*\right) - \varphi\left(\mathbf{u},t|\mathbf{x},\boldsymbol{\theta}^*\right)\right.$$
$$\left. \times\, \varphi\left(\mathbf{v},t|\mathbf{x},\boldsymbol{\theta}^*\right)\right|^2 a^2(\mathbf{x})\, d\mathbf{x} dW(\mathbf{u})\, dW(\mathbf{v})$$

$$\times \int \left[\int K(\boldsymbol{\tau}) K(\boldsymbol{\tau}+\boldsymbol{\eta}) d\boldsymbol{\tau} \right]^2 d\boldsymbol{\eta}, \qquad \varphi(\mathbf{u},t|\mathbf{x},\boldsymbol{\theta})$$

$$\equiv \varphi(\mathbf{u},t|\mathbf{X}_{t-1}=\mathbf{x},\boldsymbol{\theta}).$$

(ii) Suppose in addition that under $\mathbb{H}_A$, *the set* $S = \{(\mathbf{u},\mathbf{x}) \in \mathbb{R}^{2d} : \mathrm{E}[Z_t(\mathbf{u},\boldsymbol{\theta}^*) |\mathbf{X}_{t-1}=\mathbf{x}] \neq 0\}$ *has a strictly positive Lebesgue measure on the support* $\mathbb{W} \times \mathbf{G}$, *where* $\mathbb{W}$ *and* $\mathbf{G}$ *are the supports of weighting functions* $W(\cdot)$ *and* $a(x)$, *respectively. Furthermore, suppose that* $W(\cdot)$ *is any positive, monotonically increasing, and continuous weighting function on* $\mathbb{W}$ *and* $a(\cdot)$ *is any positive, continuous weighting function on* $\mathbf{G}$. *Then* $P[\hat{M} > C(T)] \to 1$ *as* $T \to \infty$, *for any nonstochastic sequence of constants* $\{C(T) = o(Th^{d/2})\}$.

Under the assumptions of Theorem 3(ii), $\hat{M}$ has asymptotic unit power at any given significance level, whenever S has a positive measure on support $\mathbb{W} \times \mathbf{G}$. We note that under $\mathbb{H}_A$, $\hat{M}$ diverges to infinity at the rate of $Th^{d/2}$, which is faster than the rate Th^d of a nonparametric transition density–based test (e.g., for $d = 1$, Aït-Sahalia et al., 2009; Hong and Li, 2005). It could be shown that the M test is asymptotically more powerful than a nonparametric transition density–based test in terms of the Bahadur (1960) asymptotic slope criterion, which is pertinent for power comparison under fixed alternatives.[8] Similarly, although we do not examine the asymptotic local power, we expect that $\hat{M}$ can detect a class of local alternatives converging to $\mathbb{H}_0$ at the rate of $T^{-1/2}h^{-d/4}$, whereas the transition density–based test can only detect a class of local alternatives with a slower rate of $T^{-1/2}h^{-d/2}$. This is an advantage of the nonparametric regression-based CCF testing over the nonparametric transition density approach, because of the dimension reduction. Theorem 3 remains valid if h is replaced by a data-dependent bandwidth $\hat{h}$, where $(\hat{h}-h)/h = o_P(1)$.

Unlike Chen and Hong (2005), we maintain the Markov property of $\mathbf{X}_t$ under $\mathbb{H}_A$. In this case, the $\hat{M}$ test is expected to have better power than the Chen and Hong (2005) test in detecting functional misspecification of the drift, diffusion, jump, and conditional correlation functions. In contrast, the Chen and Hong (2005) test is expected to have better power when $\mathbf{X}_t$ is not Markov under $\mathbb{H}_A$.

The finite-sample omnibus test $\hat{M}_{FS}$ in (3.8) has the same asymptotic $N(0,1)$ distribution under $\mathbb{H}_0$ and the same asymptotic power property under $\mathbb{H}_A$ as the $\hat{M}$ test.

5. DIRECTIONAL DIAGNOSTIC PROCEDURES

When a multifactor continuous-time model $\mathcal{M}$ is rejected by the omnibus test, it would be interesting to explore possible sources of the rejection. For example, one might like to know whether the misspecification comes from conditional mean dynamics, or conditional variance dynamics, or conditional correlations between state variables. Such information will be valuable in reconstructing the model.

The CCF is a convenient and useful tool to check possible sources of model misspecification. As is well known, the CCF can be differentiated to obtain conditional moments. We now develop a class of diagnostic tests in a unified framework by differentiating $m(\mathbf{u}, \mathbf{x}, \boldsymbol{\theta})$ with respect to $\mathbf{u}$ at the origin. This class of diagnostic tests can provide useful information about how well a continuous-time model captures the dynamics of various conditional moments and conditional cross-moments of state variables.

Recall the partial derivative of function $m(\mathbf{u}, \mathbf{X}_{t-1}, \boldsymbol{\theta})$ at $\mathbf{u} = \mathbf{0}$:

$$m^{(\boldsymbol{\nu})}(\mathbf{0}, \mathbf{X}_{t-1}, \boldsymbol{\theta}) = \mathrm{E}\left[\prod_{c=1}^{d}(iX_{c,t})^{\nu_c}|\mathbf{X}_{t-1}\right] - \mathrm{E}_{\boldsymbol{\theta}}\left[\prod_{c=1}^{d}(iX_{c,t})^{\nu_c}|\mathbf{X}_{t-1}\right]. \tag{5.1}$$

To get insight into $m^{(\boldsymbol{\nu})}\left(\mathbf{0}, \mathbf{X}_{t-1}, \boldsymbol{\theta}\right)$, we consider a bivariate process $\mathbf{X}_t = (X_{1,t}, X_{2,t})'$:

Case 1 ($|\boldsymbol{\nu}| = 1$). We have $\boldsymbol{\nu} = (1,0)$ or $\boldsymbol{\nu} = (0,1)$. If $\boldsymbol{\nu} = (1,0), m^{(\boldsymbol{\nu})}\left(\mathbf{0}, \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) = i\mathrm{E}\left(X_{1,t}|\mathbf{X}_{t-1}\right) - i\mathrm{E}_{\boldsymbol{\theta}}(X_{1,t}|\mathbf{X}_{t-1})$. If $\boldsymbol{\nu} = (0,1)$, then $m^{(\boldsymbol{\nu})}\left(\mathbf{0}, \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) = i\mathrm{E}\left(X_{2,t}|\mathbf{X}_{t-1}\right) - i\mathrm{E}_{\boldsymbol{\theta}}(X_{2,t}|\mathbf{X}_{t-1})$. Thus, the choice of $|\boldsymbol{\nu}| = 1$ checks misspecifications in the conditional means of $X_{1,t}$ and $X_{2,t}$, respectively.

Case 2 ($|\boldsymbol{\nu}| = 2$). We have $\boldsymbol{\nu} = (2,0), (0,2)$, or $(1,1)$. If $\boldsymbol{\nu} = (2,0)$, $m^{(\boldsymbol{\nu})}\left(\mathbf{0}, \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) = -\left[\mathrm{E}\left(X_{1,t}^2|\mathbf{X}_{t-1}\right) - \mathrm{E}_{\boldsymbol{\theta}}(X_{1,t}^2|\mathbf{X}_{t-1})\right]$. If $\boldsymbol{\nu} = (0,2)$, $m^{(\boldsymbol{\nu})}\left(\mathbf{0}, \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) = -\left[\mathrm{E}\left(X_{2,t}^2|\mathbf{X}_{t-1}\right) - \mathrm{E}_{\boldsymbol{\theta}}(X_{2,t}^2|\mathbf{X}_{t-1})\right]$. Finally, if $\boldsymbol{\nu} = (1,1)$, $m^{(\boldsymbol{\nu})}\left(\mathbf{0}, \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) = -\left[\mathrm{E}\left(X_{1,t}X_{2,t}|\mathbf{X}_{t-1}\right) - \mathrm{E}_{\boldsymbol{\theta}}(X_{1,t}X_{2,t}|\mathbf{X}_{t-1})\right]$. Thus, the choice of $|\boldsymbol{\nu}| = 2$ checks model misspecifications in the conditional volatility of state variables and the noncentered conditional covariance.

We now define the class of diagnostic test statistics as follows:

$$\hat{M}^{(\boldsymbol{\nu})} = \left[h^{d/2}\sum_{t=2}^{T}\left|\hat{m}^{(\boldsymbol{\nu})}(\mathbf{0}, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}})\right|^2 a\left(\mathbf{X}_{t-1}\right) - \hat{C}^{(\boldsymbol{\nu})}\right] \Big/ \sqrt{2\hat{D}^{(\boldsymbol{\nu})}}, \quad \text{where}$$

$$\hat{C}^{(\boldsymbol{\nu})} = h^{-d/2}\int\left\{\mathrm{E}_{\hat{\boldsymbol{\theta}}}\left(\prod_{c=1}^{d}X_{c,t}^{2\nu_c}|\mathbf{X}_{t-1} = \mathbf{x}\right) - \left[\mathrm{E}_{\hat{\boldsymbol{\theta}}}\left(\prod_{c=1}^{d}X_{c,t}^{\nu_c}|\mathbf{X}_{t-1} = \mathbf{x}\right)\right]^2\right\}$$

$$\times\, a\left(\mathbf{x}\right)d\mathbf{x}\int K^2\left(\boldsymbol{\tau}\right)d\boldsymbol{\tau},$$

$$\hat{D}^{(\boldsymbol{\nu})} = \int\left\{\mathrm{E}_{\hat{\boldsymbol{\theta}}}\left(\prod_{c=1}^{d}X_{c,t}^{2\nu_c}|\mathbf{X}_{t-1} = \mathbf{x}\right) - \left[\mathrm{E}_{\hat{\boldsymbol{\theta}}}\left(\prod_{c=1}^{d}X_{c,t}^{\nu_c}|\mathbf{X}_{t-1} = \mathbf{x}\right)\right]^2\right\}^2$$

$$\times\, a^2(\mathbf{x})d\mathbf{x}\int\left[\int K\left(\boldsymbol{\tau}\right)K\left(\boldsymbol{\tau} + \boldsymbol{\eta}\right)d\boldsymbol{\tau}\right]^2 d\eta.$$

Here, $\mathrm{E}_{\hat{\boldsymbol{\theta}}}(\cdot|\mathbf{X}_{t-1})$ is the expectation under the estimated model-implied transition density $p(\cdot,t|\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})$. In general, we can differentiate the estimated CCF $\varphi\left(\mathbf{u},t|\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}}\right)$ to obtain $\mathrm{E}_{\hat{\boldsymbol{\theta}}}(\cdot|\mathbf{X}_{t-1})$. For example,

$$\mathrm{E}_{\hat{\boldsymbol{\theta}}}\left(\prod_{c=1}^{d} X_{c,t}^{\nu_c}|\mathbf{X}_{t-1}\right) = i^{\Sigma_{c=1}^{d}\nu_c}\,\frac{\partial^{\nu_1}}{\partial u_1^{\nu_1}}\cdots\frac{\partial^{\nu_d}}{\partial u_d^{\nu_d}}\varphi\left(\mathbf{u},t\middle|\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}}\right)\Big|_{\mathbf{u}=\mathbf{0}}.$$

In the previous bivariate case, if we further assume that the DGP is the bivariate uncorrelated Gaussian model in (7.9), then for $\boldsymbol{\nu}=(1,0)$, we have

$$\begin{aligned}\mathrm{E}_{\hat{\boldsymbol{\theta}}}\left(X_{1,t}^{2\nu_1}X_{2,t}^{2\nu_2}|\mathbf{X}_{t-1}\right) &= \left\{\left[1-\exp\left(-\hat{\kappa}_{11}\right)\right]\hat{\theta}_1+\exp\left(-\hat{\kappa}_{11}\right)X_{1,t-1}\right\}^2\\ &\quad+\frac{\hat{\sigma}_{11}^2}{2\hat{\kappa}_{11}}\left[1-\exp\left(-2\hat{\kappa}_{22}\right)\right],\\ \mathrm{E}_{\hat{\boldsymbol{\theta}}}\left(X_{1,t}^{\nu_1}X_{2,t}^{\nu_2}|\mathbf{X}_{t-1}\right) &= \left[1-\exp\left(-\hat{\kappa}_{11}\right)\right]\hat{\theta}_1+\exp\left(-\hat{\kappa}_{11}\right)X_{1,t-1}.\end{aligned}$$

To derive the limit distribution of $\hat{M}^{(\boldsymbol{\nu})}$ under $\mathbb{H}_0$, we impose some moment conditions.

Assumption A.8.

(a) $\mathrm{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\left[\frac{\partial^{\nu_1}}{\partial u_1^{\nu_1}}\cdots\frac{\partial^{\nu_d}}{\partial u_d^{\nu_d}}\varphi\left(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}\right)|_{\mathbf{u}=\mathbf{0}}\right]\right\|^2 \le C$;

(b) $\mathrm{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\left[\frac{\partial^{\nu_1}}{\partial u_1^{\nu_1}}\cdots\frac{\partial^{\nu_d}}{\partial u_d^{\nu_d}}\varphi\left(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}\right)|_{\mathbf{u}=\mathbf{0}}\right]\right\|^2 \le C$;

(c) $\mathrm{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^{\nu_1}}{\partial u_1^{\nu_1}}\cdots\frac{\partial^{\nu_d}}{\partial u_d^{\nu_d}}\varphi\left(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}\right)|_{\mathbf{u}=\mathbf{0}}\right\|^{4(1+\delta)} \le C$; and

(d) $\mathrm{E}\left|\prod_{c=1}^{d}X_{c,t}^{\nu_c}\right|^{4(1+\delta)} \le C$.

THEOREM 4. *Suppose Assumptions A.1–A.5 and A.8 hold for some prespecified derivative order vector $\boldsymbol{\nu}$, $h=cT^{-\lambda}$ for $0<\lambda<\frac{1}{2d}$, and $0<c<\infty$. Then $\hat{M}^{(\boldsymbol{\nu})}\xrightarrow{d}N(0,1)$ under $\mathbb{H}_0$ as $T\to\infty$.*

Like $\hat{M}$, parameter estimation uncertainty in $\hat{\boldsymbol{\theta}}$ has no impact on the asymptotic distribution of $\hat{M}^{(\boldsymbol{\nu})}$. Any $\sqrt{T}$-consistent estimator can be used. Moreover, different choices of $\boldsymbol{\nu}$ allow one to examine various specific dynamic aspects of the underlying process and thus provide information on how well a multivariate continuous-time Markov model fits various aspects of the conditional distribution of $\mathbf{X}_t$. Theorem 4 remains valid if h is replaced by a data-dependent bandwidth $\hat{h}$, where $(\hat{h}-h)/h=o_P(h^{d/2})$.

These diagnostic tests are designed to test specification of various conditional moments, i.e., whether the conditional moments of $\mathbf{X}_t$ are correctly specified given the discrete sample information $\mathbf{X}_{t-1}$. We note that the first two conditional moments differ from the instantaneous conditional mean (drift) and instantaneous conditional variance (squared diffusion). In general, the conditional moments tested here are functions of drift, diffusion, and jump (see Section 7.1.2 for an example). Only when the sampling interval $\Delta \to 0$ will the conditional mean and variance coincide with drift and squared diffusion.[9]

6. TESTS FOR MODELS WITH UNOBSERVABLE VARIABLES

So far we have assumed that all state variables in $\mathbf{X}_t$ are observable. However, there are continuous-time models with unobservable components. For example, within the family of asset pricing models, unobserved state variables typically arise when the dimension d of $\mathbf{X}_t$ exceeds the dimension p of the vector of observed prices or yields. In the context of ATSMs, if r_t is an affine function of d state variables and one estimates the model with only p $(< d)$ bond yields, then $d - p$ remaining state variables are unobservable. Andersen and Lund (1996) estimate a three-factor model ($d = 3$) of a single short-term interest rate ($p = 1$) using the Gallant and Tauchen (1996) EMM method. Singleton (2001) also proposes a CCF-based simulated method of moments estimators to exploit the special structure of ATSMs with unobservable state variables.

Another example is the class of SV models; see, e.g., Bates (1996) and Heston (1993). With a latent volatility state variable, SV models can capture salient properties of volatility such as randomness and persistence. Affine SV models have been widely used in modeling asset return dynamics as they yield closed-form solutions for European option prices. A basic version of SV models assumes

$$\begin{cases} dr_t = \kappa_r (\bar{r} - r_t) dt + \sqrt{V_t} dW_{r,t}, \\ dV_t = \kappa_v (\bar{v} - V_t) dt + \sigma_v \sqrt{V_t} dW_{v,t}, \end{cases} \tag{6.1}$$

where V_t is the latent volatility process and $\kappa_r, \kappa_v, \sigma_v, \bar{r}$, and $\bar{v}$ are all scalar parameters. It can be shown that the CCF of r_t given (r_{t-1}, V_{t-1}) is

$$\begin{aligned} &\varphi_r \left(u, t | r_{t-1}, V_{t-1}, \boldsymbol{\theta}\right) \\ &\quad = \exp\left[A_{t-1}(u, 0) + B_{t-1}(u, 0) r_{t-1} + C_{t-1}(u, 0) V_{t-1}\right], \quad u \in \mathbb{R}, \end{aligned} \tag{6.2}$$

where $A_{t-1}, B_{t-1}, C_{t-1} : \mathbb{R}^2 \to \mathbb{R}$ satisfy the complex-valued Riccati equations:

$$\begin{cases} \dot{A}_t = \kappa_r \bar{r} B_t + \kappa_v \bar{v} C_t, \\ \dot{B}_t = -\kappa_r B_t, \\ \dot{C}_t = -\kappa_v C_t + \frac{1}{2}\left(B_t^2 + 2 B_t C_t \sigma_v + C_t \sigma_v^2\right). \end{cases}$$

To test SV models, where V_t is a latent process, we need to modify the MDS characterization (2.11).

Generally, we partition $\mathbf{X}_t = (\mathbf{X}_{1,t}', \mathbf{X}_{2,t}')'$, where $\mathbf{X}_{1,t} \subset \mathbb{R}^{d_1}$ denotes observable state variables, $\mathbf{X}_{2,t} \subset \mathbb{R}^{d_2}$ denotes unobservable state variables, and $d_1 + d_2 = d$. Also, partition $\mathbf{u}$ conformably as $\mathbf{u} = \left(\mathbf{u}_1', \mathbf{u}_2'\right)'$. Let

$$\begin{aligned}\phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}) &\equiv \mathrm{E}_{\boldsymbol{\theta}}[\exp(i\mathbf{u}_1'\mathbf{X}_{1,t})|\mathcal{I}_{1,t-1}] \\ &= \mathrm{E}_{\boldsymbol{\theta}}\left\{\varphi\left[(\mathbf{u}_1', \mathbf{0}')', t|\mathbf{X}_{t-1}, \boldsymbol{\theta}\right]|\mathcal{I}_{1,t-1}\right\},\end{aligned}$$

where the second equality follows from the law of iterated expectations and the Markov property of $\mathbf{X}_t$ and $\mathcal{I}_{1,t-1} = \left\{\mathbf{X}_{1,t-1}, \mathbf{X}_{1,t-2}, \ldots, \mathbf{X}_{1,1}\right\}$ is the information set on the observables available at time $t-1$. We define

$$Z_{1,t}(\mathbf{u}_1, \boldsymbol{\theta}) \equiv \exp\left(i\mathbf{u}_1'\mathbf{X}_{1,t}\right) - \phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}).$$

Under $\mathbb{H}_0$, we have

$$\mathrm{E}\left[Z_{1,t}(\mathbf{u}_1, \boldsymbol{\theta}_0)|\mathcal{I}_{1,t-1}\right] = 0 \text{ a.s.} \quad \text{for all } \mathbf{u}_1 \in \mathbb{R}^{d_1} \quad \text{and some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}. \tag{6.3}$$

This provides a basis for constructing operational tests for continuous-time Markov models with partially observable state variables.[10] It has been used in Singleton (2001) to estimate continuous-time models with unobservable components. Note that although $Z_t(\mathbf{u}, \boldsymbol{\theta}_0)$ is a Markov process, $Z_{1,t}(\mathbf{u}_1, \boldsymbol{\theta}_0)$ is generally not.[11]

For notational simplicity, we define a new vector $\mathbf{Y}_t = (\mathbf{X}_{1,t}', \mathbf{X}_{1,t-1}', \ldots, \mathbf{X}_{1,t-l+1}')' \subset \mathbb{R}^{ld_1}$, where l is a lag truncation order. Based on the MDS characterization in (6.3), we can use a nonparametric estimator for $m_u(\mathbf{u}_1, \mathbf{Y}_{t-1}, \boldsymbol{\theta}_0) \equiv \mathrm{E}[Z_{1,t}(\mathbf{u}_1, \boldsymbol{\theta}_0)|\mathbf{Y}_{t-1}]$. Similar to (3.1), we consider the following local least squares problem:

$$\min_{\boldsymbol{\beta}} \sum_{t=l+1}^{T} \left|Z_{1,t}(\mathbf{u}_1, \boldsymbol{\theta}_0) - \beta_0 - \boldsymbol{\beta}_1'(\mathbf{Y}_t - \mathbf{y})\right|^2 K_h(\mathbf{y} - \mathbf{Y}_t), \qquad \mathbf{y} \in \mathbb{R}^{ld_1}, \tag{6.4}$$

where $\boldsymbol{\beta} = \left(\beta_0, \boldsymbol{\beta}_1'\right)'$. We obtain the following solution:

$$\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\mathbf{y}, \mathbf{u}_1) = \begin{bmatrix} \hat{\beta}_0(\mathbf{y}, \mathbf{u}_1) \\ \hat{\boldsymbol{\beta}}_1'(\mathbf{y}, \mathbf{u}_1) \end{bmatrix} = [\mathbf{Y}'\mathbf{W}\mathbf{Y}]^{-1}\mathbf{Y}'\mathbf{W}\mathbf{Z}_1, \tag{6.5}$$

where $\mathbf{Y}$ is a $Tld_1 \times 2$ matrix with the $(t+1)$ to $(t+d)$th row $[\mathbf{1}, \mathbf{Y}_t - \mathbf{y}]$, $\mathbf{W}_y = \mathrm{diag}[K_h(\mathbf{Y}_1 - \mathbf{y}), K_h(\mathbf{Y}_2 - \mathbf{y}), \ldots, K_h(\mathbf{Y}_T - \mathbf{y})]$, and $\mathbf{Z}_1 = [Z_{1,1}(\mathbf{u}_1, \boldsymbol{\theta}_0), \ldots, Z_{1,T}(\mathbf{u}_1, \boldsymbol{\theta}_0)]'$. The function $m_u(\mathbf{u}, \mathbf{y}, \boldsymbol{\theta}_0)$ can be consistently estimated by $\hat{\beta}_0\left(\mathbf{y}, \mathbf{u}_1\right)$, namely,

$$\hat{m}_u\left(\mathbf{u}_1, \mathbf{y}, \hat{\boldsymbol{\theta}}\right) = \sum_{t=l+1}^{T} \hat{W}\left(\frac{\mathbf{Y}_t - \mathbf{y}}{h}\right) Z_{1,t}\left(\mathbf{u}_1, \hat{\boldsymbol{\theta}}\right), \qquad \mathbf{y} \in \mathbb{R}^{ld_1}, \tag{6.6}$$

where $\hat{W}$ is defined in the same way as in (3.4). The omnibus test statistic for $\mathbb{H}_0$ against $\mathbb{H}_A$ is a modified version of (3.7), namely,

$$\hat{M}_u = \left[h^{ld_1/2} \sum_{t=l+1}^{T} \int \left| \hat{m}_u \left(\mathbf{u}_1, \mathbf{Y}_{t-1z}, \hat{\boldsymbol{\theta}} \right) \right|^2 \right.$$

$$\left. \times a\left(\mathbf{Y}_{t-1}\right) dW\left(\mathbf{u}_1\right) - \hat{C}_u \right] \Big/ \sqrt{2\hat{D}_u}, \quad \text{where}$$

$$\hat{C}_u = h^{-ld_1/2} \iint \left\{ 1 - \mathrm{E}_{\hat{\boldsymbol{\theta}}} \left[\left| \phi\left(\mathbf{u}_1, t | \mathcal{I}_{1,t-1}, \hat{\boldsymbol{\theta}} \right) \right|^2 \middle| \mathbf{Y}_{t-1} = \mathbf{y} \right] \right\}$$

$$\times a(\mathbf{y})\, d\mathbf{y}\, dW\left(\mathbf{u}_1\right) \int K^2(\boldsymbol{\tau})\, d\boldsymbol{\tau},$$

$$\hat{D}_u = \iiint \left| \mathrm{E}_{\hat{\boldsymbol{\theta}}} \left[\phi\left(\mathbf{u}_1 + \mathbf{v}_1, t | \mathcal{I}_{1,t-1}, \hat{\boldsymbol{\theta}} \right) - \phi\left(\mathbf{u}_1, t | \mathcal{I}_{1,t-1}, \hat{\boldsymbol{\theta}} \right) \right.\right.$$

$$\left.\left. \times \phi\left(\mathbf{v}_1, t | \mathcal{I}_{1,t-1}, \hat{\boldsymbol{\theta}} \right) \middle| \mathbf{Y}_{t-1} = \mathbf{y} \right] \right|^2 a^2(\mathbf{y})\, d\mathbf{y}\, dW\left(\mathbf{u}_1\right) dW\left(\mathbf{v}_1\right)$$

$$\times \int \left[\int K(\boldsymbol{\tau})\, K(\boldsymbol{\tau} + \boldsymbol{\eta})\, d\boldsymbol{\tau} \right]^2 d\boldsymbol{\eta}, \tag{6.7}$$

where $W : \mathbb{R}^{ld_1} \to \mathbb{R}^+$ is a nondecreasing weighting function that weighs sets symmetric about the origin equally, $a : \mathbf{F} \to \mathbb{R}^+$ is a bounded weighting function, and $\mathbf{F} \in \mathbb{R}^{ld_1}$ is a compact support. The conditional expectations $\mathrm{E}_{\hat{\boldsymbol{\theta}}}(\cdot|\mathbf{Y}_{t-1})$ in $\hat{C}_u$ and $\hat{D}_u$ can be estimated via a nonparametric regression, but its implementation may be tedious. Alternatively, we can consider the following finite-sample version of the test statistic:

$$\hat{M}_u^{FS} = \left[h^{ld_1/2} \sum_{t=l+1}^{T} \int \left| \hat{m}_u(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}}) \right|^2 \right.$$

$$\left. \times a\left(\mathbf{Y}_{t-1}\right) dW\left(\mathbf{u}_1\right) - \hat{C}_u^{FS} \right] \Big/ \sqrt{2\hat{D}_u^{FS}}, \quad \text{where}$$

$$\hat{C}_u^{FS} = h^{ld_1/2} \sum_{s=l+1}^{T} \int \left| \mathbf{Z}_{1,s}\left(\mathbf{u}_1, \hat{\boldsymbol{\theta}} \right) \right|^2 dW\left(\mathbf{u}_1\right)$$

$$\times \sum_{t=l+1}^{T} \hat{W}^2 \left(\frac{\mathbf{Y}_{s-1} - \mathbf{Y}_{t-1}}{h} \right) a\left(\mathbf{Y}_{t-1}\right),$$

$$\hat{D}_u^{FS} = h^{ld_1/2} \sum_{s=l+1}^{T} \sum_{r=l+1}^{s-1} \left\{ \sum_{t=l+1}^{T} \int \mathrm{Re}\left[\mathbf{Z}_{1,s}\left(\mathbf{u}_1, \hat{\boldsymbol{\theta}}\right) \mathbf{Z}_{1,r}^{*}\left(\mathbf{u}_1, \hat{\boldsymbol{\theta}}\right) \right] dW(\mathbf{u}_1) \right.$$
$$\left. \times \hat{W}\left(\frac{\mathbf{Y}_{s-1} - \mathbf{Y}_{t-1}}{h}\right) \hat{W}\left(\frac{\mathbf{Y}_{r-1} - \mathbf{Y}_{t-1}}{h}\right) a(\mathbf{Y}_{t-1}) \right\}^2. \tag{6.8}$$

Both $\hat{C}_u^{FS}$ and $\hat{D}_u^{FS}$ are computationally simpler than $\hat{C}_u$ and $\hat{D}_u$ in (6.7). Moreover, similar to (3.8), $\hat{C}_u^{FS}$ and $\hat{D}_u^{FS}$ are expected to give better approximation for the mean and variance of $h^{ld_1/2} \sum_{t=l+1}^{T} \int \left|\hat{m}_u(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}})\right|^2 a\left(\mathbf{Y}_{t-1}\right) dW\left(\mathbf{u}_1\right)$ in finite samples. Consequently, $\hat{M}_u^{FS}$ is expected to deliver better sizes in finite samples.

We now examine the asymptotic behavior of $\hat{M}_u$ under $\mathbb{H}_0$ and $\mathbb{H}_A$, respectively.

THEOREM 5. *Suppose Assumptions B.1–B.6 given in the Appendix hold and* $h = cT^{-\lambda}$ *for* $0 < \lambda < \frac{1}{2ld_1}$ *and* $0 < c < \infty$. *Then* $\hat{M}_u \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0$ *as* $T \to \infty$.

Assumptions B.1–B.5 are straightforward modifications of Assumptions A.1–A.5. We impose Assumption B.6, which assumes that the CCF estimator $\hat{\phi}\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)$ converges to $\phi\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)$ at a $\sqrt{T}$-convergence rate. This is a high-level assumption, covering various consistent estimators for $\phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$. We discuss several popular methods here to estimate $\phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$. We first consider particle filters, which have been developed by Gordon, Salmond, and Smith (1993), Pitt and Shephard (1999), and Johannes, Polson, and Stroud (2009). For continuous-time Markov models, the CCF $\varphi\left(\mathbf{u}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}\right)$ is a function of $\mathbf{X}_{t-1}$. It follows that

$$\phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}) = \int \varphi[(\mathbf{u}_1', \mathbf{0}')', t|\mathbf{X}_{t-1}, \boldsymbol{\theta}] p\left[\mathbf{x}_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right] d\mathbf{x}_{2,t-1},$$

where $p[\mathbf{x}_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}]$ is the model-implied transition density of the unobservable $\mathbf{X}_{2,t-1}$ given the past observable information $\mathcal{I}_{1,t-1}$. Gordon et al. (1993) and Pitt and Shephard (1999) develop a general method called particle filters that can approximate the conditional density $p(\mathbf{x}_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$ by a set of particles $\{\hat{\mathbf{X}}_{2,t-1}^{j}\}_{j=1}^{J}$ with discrete probability masses $\{\pi_{t-1}^{j}\}_{j=1}^{J}$ for a large integer $J >> T$. The key is to propagate particles $\{\hat{\mathbf{X}}_{2,t-2}^{j}\}_{j=1}^{J}$ one step forward to get the new particles $\{\hat{\mathbf{X}}_{2,t-1}^{j}\}_{j=1}^{J}$. By the Bayes rule, we have

$$p\left(\mathbf{x}_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)$$
$$= \frac{p\left(\mathbf{x}_{1,t-1}, t-1|\mathbf{x}_{2,t-1}, \mathcal{I}_{1,t-2}, \boldsymbol{\theta}\right) p\left(\mathbf{x}_{2,t-1}, t-1|\mathcal{I}_{1,t-2}, \boldsymbol{\theta}\right)}{p\left(\mathbf{x}_{1,t-1}, t-1|\mathcal{I}_{1,t-2}, \boldsymbol{\theta}\right)},$$

where $p\left(\mathbf{x}_{2,t-1}, t-1|\mathcal{I}_{1,t-2}, \boldsymbol{\theta}\right) = \int p\left(\mathbf{x}_{2,t-1}, t-1|\mathbf{x}_{2,t-2}, \mathcal{I}_{1,t-2}, \boldsymbol{\theta}\right) p(\mathbf{x}_{2,t-2}, t-2|\mathcal{I}_{1,t-2}, \boldsymbol{\theta}) d\mathbf{x}_{2,t-2}$. We can then approximate $p\left(\mathbf{x}_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)$ up to some proportionality, namely,

$$\hat{p}\left(\mathbf{x}_{2,t-1}, t-1|\hat{\mathcal{I}}_{1,t-1}, \boldsymbol{\theta}\right) \propto \hat{p}\left(\mathbf{x}_{1,t-1}, t-1|\hat{\mathbf{X}}_{2,t-1}, \hat{\mathcal{I}}_{1,t-2}, \boldsymbol{\theta}\right) \times \sum_{j=1}^{J} \pi_{t-1}^{j} \hat{p}\left(\mathbf{x}_{2,t-1}, t-1|\hat{\mathbf{X}}_{2,t-2}, \hat{\mathcal{I}}_{1,t-2}, \boldsymbol{\theta}\right),$$

where $\hat{\mathcal{I}}_{1,t-1} = \{\hat{\mathbf{X}}_{1,t-1}, \hat{\mathbf{X}}_{1,t-2}, \ldots, \hat{\mathbf{X}}_{1,1}\}$ and $\hat{p}(\mathbf{x}_{1,t-1}, t-1|\hat{\mathbf{X}}_{2,t-1}, \hat{\mathcal{I}}_{1,t-2}, \boldsymbol{\theta})$ and $\sum_{j=1}^{J} \pi_{t-1}^{j} \hat{p}(\mathbf{x}_{2,t-1}, t-1|\hat{\mathbf{X}}_{2,t-2}, \hat{\mathcal{I}}_{1,t-2}, \boldsymbol{\theta})$ can be viewed as the likelihood and prior, respectively. As pointed out by Gordon et al. (1993), the particle filters require that the likelihood function can be evaluated and that $\hat{\mathbf{X}}_{2,t-1}$ can be sampled from $\hat{p}(\mathbf{x}_{2,t-1}, t-1|\hat{\mathbf{X}}_{2,t-2}, \hat{\mathcal{I}}_{1,t-2}, \boldsymbol{\theta})$. These can be achieved by using the Euler or Milstein scheme (e.g., Kloeden et al., 1994).

To implement particle filters, one can follow the algorithm developed by Johannes et al. (2009).[12] First we generate a simulated sample $\{(\hat{\mathbf{X}}_{2,t-2}^{M})^{j}\}_{j=1}^{J}$, where $\hat{\mathbf{X}}_{2,t-2}^{M} = \left\{\hat{\mathbf{X}}_{2,t-2\Delta}, \hat{\mathbf{X}}_{2,t-2\Delta+\frac{\Delta}{M}}, \ldots, \hat{\mathbf{X}}_{2,t-2\Delta+\frac{(M-1)\Delta}{M}}\right\}$ and M is the number of intermediate steps between observations $\hat{\mathbf{X}}_{2,t-2\Delta}$ and $\hat{\mathbf{X}}_{2,t-\Delta}$. Then we simulate them one step forward, evaluate the likelihood function, and set

$$\pi_{t-1}^{j} = \frac{\hat{p}[\mathbf{x}_{1,t-1}, t-1|(\hat{\mathbf{X}}_{2,t-1}^{M})^{j}, \hat{\mathcal{I}}_{1,t-2}, \boldsymbol{\theta}]}{\sum_{j=1}^{J} \hat{p}[\mathbf{x}_{1,t-1}, t-1|(\hat{\mathbf{X}}_{2,t-1}^{M})^{j}, \hat{\mathcal{I}}_{1,t-2}, \boldsymbol{\theta}]}, \qquad j = 1, \ldots, J.$$

Finally, we resample J particles with weights $\{\pi_{t-1}^{j}\}_{j=1}^{J}$ to obtain a new random sample of size J.[13] It has been shown (e.g., Bally and Talay, 1996; Del Moral, Jacod, and Protter, 2001) that with both large J and M, this algorithm sequentially generates valid simulated samples from $p(\mathbf{x}_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$. Hence $\hat{\phi}(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$ can be calculated by Monte Carlo averages.

Asymptotic convergence results for diffusion models are provided by Del Moral and Jacod (2001) and Del Moral et al. (2001). They combine the Bally and Talay (1996) result of pointwise convergence with standard particle convergence to show that the particle filter is consistent provided that M increases move slowly than J and the convergence rate is proportional to $J^{-1/(3+d_1)}$. Therefore, as long as $J > T^{(3+d_1)/2}$, our assumption on the convergence rate of the estimator $\hat{\phi}(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$ is expected to hold. Asymptotic properties of the particle filter for jump diffusions remain an open question, although Hausenblas (2002) has conjectured that the Bally and Talay (1996) approach can be extended to jump diffusions. Nevertheless, Johannes et al. (2009) show that accurate estimates can be obtained in jump diffusion models via extensive simulations.

The second method to approximate $p(\mathbf{x}_{2,t-1}, t-1|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$ is the Gallant and Tauchen (1998) seminonparametric (SNP) reprojection technique, which can characterize the dynamic response of a partially observed nonlinear system to its past observable history. First, we can generate simulated samples $\{\hat{\mathbf{X}}_{1,t-1}\}_{t=2}^{J}$ and $\{\hat{\mathbf{X}}_{2,t-1}\}_{t=2}^{J}$ from the continuous-time model, where J is a large integer. Then, we project the simulated data $\{\hat{\mathbf{X}}_{2,t-1}\}_{t=2}^{J}$ onto a Hermite series representation of the transition density $p(\mathbf{x}_{2,t-1},t-1|\hat{\mathbf{X}}_{1,t-1},\hat{\mathbf{X}}_{1,t-2},\ldots\hat{\mathbf{X}}_{1,t-L})$, where L denotes a truncation lag order. With a suitable choice of L via some information criteria such as the Akaike information criterion or Bayesian information criterion, we can approximate $p(\mathbf{x}_{2,t-1},t-1|\hat{\mathcal{I}}_{1,t-1},\boldsymbol{\theta})$ arbitrarily well. The final step is to evaluate the estimated density function at the observed data in the conditional information set. See Gallant and Tauchen (1998) for more discussion. Fenton and Gallant (1996) show that the convergence rate of the SNP density estimator is $\left(J^{-1/2+\alpha/2+\delta}\right)$ with the L_1 norm, where δ is some small number and $0<\alpha<1$, which is $O_P(T^{-1/2})$ if $J>T^{1/(1-\alpha-2\delta)}$. We conjecture a similar result for our convergence rate assumption on $\hat{\phi}(\mathbf{u}_1,t|\mathcal{I}_{1,t-1},\boldsymbol{\theta})$.

For models whose CCF is exponentially affine in $\mathbf{X}_{t-1}$,[14] we can also adopt the Bates (2007) approach to compute $\hat{\phi}(\mathbf{u}_1,t|\mathcal{I}_{1,t-1},\boldsymbol{\theta})$. First, at time $t=1$, we initialize the CCF of the latent vector $\mathbf{X}_{2,t-1}$ conditional on $\mathcal{I}_{1,t-1}$ at its unconditional characteristic function. Then, by exploiting the Markov property and the affine structure of the CCF, we can evaluate the model-implied CCF conditional on data observed through period t, namely, $\mathrm{E}_{\boldsymbol{\theta}}[\varphi(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta})|\mathcal{I}_{1,t-1}]$, and thus an estimator for $\phi(\mathbf{u}_1,t|\mathcal{I}_{1,t-1},\boldsymbol{\theta})$ is obtained.

Like $\hat{M}$, the use of the estimated generalized residuals $\{Z_{1,t}(\mathbf{u}_1,\hat{\boldsymbol{\theta}})\}$ in place of the true unobservable residuals $\left\{Z_{1,t}\left(\mathbf{u}_1,\boldsymbol{\theta}_0\right)\right\}$ has no impact on the limit distribution of $\hat{M}_u$. One can proceed as if the true parameter value $\boldsymbol{\theta}_0$ were known and equal to $\hat{\boldsymbol{\theta}}$. Theorem 5 remains valid if h is replaced by a data-dependent bandwidth $\hat{h}$, where $(\hat{h}-h)/h=o_P(h^{ld_1/2})$. We do not state an additional theorem here to save space.

Next we consider the asymptotic behavior of $\hat{M}_u$ under $\mathbb{H}_A$.

THEOREM 6. *Suppose Assumptions B.1, B.2, B.4, B.5, B.7, and B.8 hold and* $h=cT^{-\lambda}$ *for* $0<\lambda<\frac{2}{3ld_1}$ *and* $0<c<\infty$. *Then* $T^{-1}h^{-ld_1/2}\hat{M}_u \xrightarrow{p} (2D_u)^{-1/2}\iint |m_u(\mathbf{u}_1,\mathbf{y},\boldsymbol{\theta}^*)|^2 a(\mathbf{y}) f(\mathbf{y})\,d\mathbf{y}\,dW(\mathbf{u}_1)$ *as* $T\to\infty$, *where* $f(\cdot)$ *is the stationary probability density function of* $\mathbf{Y}_t$ *and the scaling factor*

$$D_u=\iiint \Big|\mathrm{E}_{\theta^*}\left[\phi\left(\mathbf{u}_1+\mathbf{v}_1,t|\mathcal{I}_{1,t-1},\boldsymbol{\theta}^*\right)-\phi\left(\mathbf{u}_1,t|\mathcal{I}_{1,t-1},\boldsymbol{\theta}^*\right)\right.$$
$$\left.\times\phi\left(\mathbf{v}_1,t|\mathcal{I}_{1,t-1},\boldsymbol{\theta}^*\right)|\mathbf{y}\right]\Big|^2 a^2(\mathbf{y})\,d\mathbf{y}\,dW(\mathbf{u}_1)\,dW(\mathbf{v}_1)$$
$$\times\int\left[\int K(\boldsymbol{\tau})K(\boldsymbol{\tau}+\boldsymbol{\eta})\,d\boldsymbol{\tau}\right]^2 d\boldsymbol{\eta}.$$

Equation (6.3) is a necessary but not necessarily sufficient condition for correct model specification with unobservable components. It is possible that equation (6.3) holds but the model is misspecified. In this case, our test $\hat{M}_u$ will have no power. However, this is not particular to our test (see, e.g., Bhardwaj et al., 2008), because only the observable $\mathbf{X}_{1,t}$ is available and the unobservable $\mathbf{X}_{2,t}$ has to be integrated out, which generally results in some loss of information (and so some loss of power). Theorem 6 remains valid if h is replaced by a data-dependent bandwidth $\hat{h}$, where $(\hat{h}-h)/h = o_P(1)$. We do not state an additional theorem here to save space.

The finite-sample version of the test statistic $\hat{M}_u^{FS}$ in (6.8) has the same asymptotic $N(0,1)$ distribution under $\mathbb{H}_0$ and the same asymptotic power property under $\mathbb{H}_A$ as the $\hat{M}_u$ test.

Like other nonparametric tests in time domain, the test $\hat{M}_u$ is subject to the well-known "curse of dimensionality" when the truncation lag order l is large. However, as in a related context (e.g., Skaug and Tjøstheim, 1996), we can use a pairwise testing approach and consider the following alternative test statistic: $\hat{M}_u = \sum_{j=1}^{l} \hat{M}_u(j)$, where $\hat{M}_u(j)$ is an appropriately centered and scaled version of the statistic $h^{d_1/2}\sum_{t=l+1}^{T}\int \left|\hat{m}_u(\mathbf{u}_1, \mathbf{X}_{1,t-j}, \hat{\boldsymbol{\theta}})\right|^2 a\left(\mathbf{X}_{1,t-j}\right) dW(\mathbf{u}_1)$. This avoids the "curse of dimensionality" with nonparametric estimation. By a similar but more tedious proof, we could derive the asymptotic normality of $\hat{M}_u$ under $\mathbb{H}_0$. Alternatively, a parametric bootstrap can be used, which in fact gives better size in finite samples.

7. MONTE CARLO EVIDENCE

We now study the finite-sample performance of the proposed tests, in comparison with the Hong and Li (2005) test. We consider both univariate and bivariate continuous-time models.

7.1. Univariate Models

7.1.1. Size of the $\hat{M}$ Tests. To examine the size of $\hat{M}$ for univariate continuous-time Markov models, we simulate data from the Vasicek (1977) model (DGP A0):

$$dX_t = \kappa(\alpha - X_t)\,dt + \sigma\,dW_t, \qquad \textbf{(7.1)}$$

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 the slower the convergence to the long-run mean. We are particularly interested in the possible impact of dependent persistence in $\{X_t\}$ on the size of $\hat{M}$. Because the finite-sample performance of $\hat{M}$ may depend on both the marginal density and dependent persistence of $\{X_t\}$, we follow Hong and Li (2005) and Pritsker (1998) to change κ and σ^2 in the same proportion so that the marginal density of X_t 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 $\boldsymbol{\theta} = (\kappa,\alpha,\sigma^2)'$ and $\sigma_s^2 = \sigma^2/(2\kappa) = 0.01226$. In this way, we can focus on the impact of dependent persistence. We consider both low and high levels of dependent persistence and use the same parameter values as Hong and Li (2005) and Pritsker (1998): $(\kappa, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$ and $(0.214592, 0.089102, 0.000546)$ for the low and high persistent dependence cases, respectively.

We simulate 1,000 data sets of a random sample $\{X_t\}_{t=\Delta}^{T\Delta}$ at the monthly frequency ($\Delta = \frac{1}{12}$) for $T = 250, 500, 1{,}000$, respectively.[15] We first generate an initial value X_0 from the marginal density $p(x, \boldsymbol{\theta}_0)$ and then, given a value X_t, generate X_{t+1} from the transitional normal density with mean and variance

$$\mu_t = X_t \exp(-\kappa\Delta) + \alpha[1 - \exp(-\kappa\Delta)], \tag{7.2}$$

$$\sigma_t^2 = \sigma^2[1 - \exp(-2\kappa\Delta)]/(2\kappa). \tag{7.3}$$

The sample sizes of $T = 250, 500, 1{,}000$ correspond to about 20–100 years of monthly data. For each data set, we estimate a Vasicek model via MLE and compute the $\hat{M}$ statistic. We consider the empirical rejection rates using the asymptotic critical values at the 10% and 5% significance levels, respectively. For $T = 250$, we also consider a parametric bootstrap procedure.

Following Aït-Sahalia et al. (2001), we use the Gaussian kernel $k(\cdot)$ and the truncated weighting $a(x) = \mathbf{1}(|x| \le 1.5)$, where $\mathbf{1}(\cdot)$ is the indicator function and X_t has been standardized to have unit variance. We choose the $N(0,1)$ cdf for $W(\cdot)$. Our simulation experience suggests that the choices of $k(\cdot)$, $W(\cdot)$, and $a(\cdot)$ have little impact on the size performance of the tests. For simplicity, we choose $h = T^{-\frac{1}{5}}$. This simple bandwidth rule attains the optimal rate for the local linear regression fitting.

Table 1 reports the empirical sizes of $\hat{M}$ and $\hat{M}^{(\nu)}$ at the 10% and 5% levels under a correct Vasicek model with low and high persistence of dependence, respectively. Both the asymptotic version in (3.7) and the finite-sample version in (3.8) of our omnibus test tend to overreject when $T = 250$, but they improve as T increases. As expected, the finite-sample version $\hat{M}_{FS}$ has better sizes. The tests display a bit more overrejection under high persistence than under low persistence, but the difference becomes smaller as T increases. For comparison, Table 2 reports the empirical sizes of the Hong and Li (2005) test under the same DGPs. Similarly, the Hong and Li test has some overrejection that is a bit more severe than that of the $\hat{M}$ tests. We also consider the diagnostic tests $\hat{M}^{(\nu)}$ for $\nu = 1, 2$, which check model misspecifications in the conditional mean and conditional variance of the state variable. The $\hat{M}^{(\nu)}$ tests have similar size patterns as $\hat{M}$ except that the overrejection is more severe in small samples.

Because the sizes of our tests using asymptotic theory differ significantly from the asymptotic significance level in small samples, we also consider the following parametric bootstrap procedure.

Step (i). Use, e.g., the Euler scheme or the generalized Milstein scheme to obtain a bootstrap sample $\mathcal{X}^b \equiv \{X_t^b\}_{t=\Delta}^{T\Delta}$ from the estimated null model

$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t, \hat{\boldsymbol{\theta}})dt + \boldsymbol{\sigma}(\mathbf{X}_t, \hat{\boldsymbol{\theta}})d\mathbf{W}_t + dJ_t(\hat{\boldsymbol{\theta}})$. To obtain monthly data ($\Delta = \frac{1}{12}$), we generate $M = 120$ intermediate steps between $t - \Delta$ and t.

Step (ii). Estimate the null model using the bootstrap sample $\mathcal{X}^b$ and compute a bootstrap statistic $\hat{M}^b$ in the same way as $\hat{M}$, with $\mathcal{X}^b$ replacing the original sample $\mathcal{X} = \{X_t\}_{t=\Delta}^{T\Delta}$.

Step (iii). Repeat steps (i) and (ii) B times to obtain B bootstrap test statistics $\{\hat{M}_l^b\}_{l=1}^B$.[16]

Step (iv). Compute the bootstrap P-value $P_b \equiv B^{-1}\sum_{l=1}^B \mathbf{1}(\hat{M}_l^b > \hat{M})$. To obtain an accurate bootstrap P-value, B must be sufficiently large.

The parametric bootstrap has been widely used to improve finite-sample performance of model specification tests. For example, Fan, Li, and Min (2006) and Li and Tkacz (2006) apply it to testing for correct specification of parametric conditional distribution and conditional density, respectively; Aït-Sahalia et al. (2009) use it to evaluate jump diffusion models. We can show that conditional on $\mathcal{X}$, $\hat{M}^b \xrightarrow{d} N(0, 1)$ in probability as $T \to \infty$.[17] Under $\mathbb{H}_0$, the bootstrap procedure will lead to asymptotically correct size of the test, because $\hat{M}^b$ converges in distribution to $N(0, 1)$; when $\mathbb{H}_0$ is false, $\hat{M} \to \infty$ in probability as $T \to \infty$, but the bootstrap critical value is still the same as that of $N(0, 1)$. As a result, the bootstrap procedure has power.

The consistency of the parametric bootstrap does not indicate the degree of improvement of the parametric bootstrap upon the asymptotic distribution. Because $\hat{M}$ is asymptotically pivotal, it is possible that $\hat{M}^b$ can achieve reasonable accuracy in finite samples. Because of the high computational cost in simulation studies, we only consider bootstraps for $T = 250$. We generate 500 data sets of random sample $\{X_t\}_{t=\Delta}^{T\Delta}$ and use $B = 100$ bootstrap iterations for each simulated data set. Table 1 shows that the bootstrap indeed approximates the finite-sample distribution of test statistics more accurately. In particular, the bootstrap significantly reduces the overrejection of the asymptotic version of our derivative tests. The bootstrap improvement of the finite-sample version $\hat{M}_{FS}$ is less significant because $\hat{M}_{FS}$ has achieved reasonable sizes using the asymptotic theory.

7.1.2. Power of the $\hat{M}$ Tests. To examine the power of $\hat{M}$ in differentiating a Vasicek model from other alternatives, we simulate data from five continuous-time models and test the null hypothesis that data are generated from a Vasicek model. The first four models and the last model have been considered in Hong and Li (2005) and Aït-Sahalia et al. (2009), respectively:

DGP A1 [Cox, Ingersoll, and Ross (CIR, 1985) model].

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t, \tag{7.4}$$

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

TABLE 1. Sizes of specification tests under DGP A0

	Significance level							
	.10		.05		.10		.05	
	BCV-AS	BCV-FS	BCV-AS	BCV-FS	ACV-AS	ACV-FS	ACV-AS	ACV-FS
	$T = 250$				$T = 250$			
Low persistence								
$\hat{M}$	.126	.128	.064	.068	.159	.108	.100	.064
$\hat{M}^{(1)}$	.134	.140	.072	.068	.166	.107	.106	.079
$\hat{M}^{(2)}$	.110	.104	.046	.052	.165	.123	.126	.095
High persistence								
$\hat{M}$	.112	.112	.056	.058	.191	.127	.129	.087
$\hat{M}^{(1)}$	.116	.114	.058	.052	.209	.146	.141	.100
$\hat{M}^{(2)}$	.098	.094	.052	.060	.238	.189	.193	.156
	ACV-AS	ACV-FS	ACV-AS	ACV-FS	ACV-AS	ACV-FS	ACV-AS	ACV-FS
	$T = 500$				$T = 1,000$			
Low persistence								
$\hat{M}$	.136	.098	.085	.055	.132	.104	.086	.067
$\hat{M}^{(1)}$	.142	.113	.100	.074	.151	.120	.100	.077
$\hat{M}^{(2)}$	.139	.099	.089	.063	.137	.097	.085	.061
High persistence								
$\hat{M}$	.136	.100	.082	.062	.144	.106	.098	.079
$\hat{M}^{(1)}$	.157	.124	.115	.087	.153	.124	.111	.091
$\hat{M}^{(2)}$	.185	.148	.145	.111	.158	.118	.108	.084

Note: DGP A0 is the Vasicek (1977) model, given in equation (7.1). Low persistence and high persistence correspond to $(K, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$ and (0.214592, 0.089102, 0.000546), respectively; $\hat{M}, \hat{M}^{(1)}$, and $\hat{M}^{(2)}$ are the omnibus test, the conditional mean test, and the conditional variance test, respectively; BCV-AS and BCV-FS denote the asymptotic version and the finite-sample version using bootstrap critical values, respectively; ACV-AS and ACV-FS denote the asymptotic version and the finite-sample version using asymptotic critical values, respectively. The p values of ACV-AS and ACV-FS are based on the results of 1,000 iterations; the p values of BCV-AS and BCV-FS are based on the results of 500 iterations.

DGP A2 (Ahn and Gao, 1999, inverse-Feller model).

$$dX_t = X_t \left[\kappa - \left(\sigma^2 - \kappa\alpha\right) X_t\right] dt + \sigma X_t^{3/2} dW_t, \tag{7.5}$$

where $(\kappa, \alpha, \sigma^2) = (3.4387, 0.0828, 1.420864)$.[18]

DGP A3 (Chan, Karolyi, Longstaff, and Sanders, 1992, model).

$$dX_t = \kappa(\alpha - X_t)\, dt + \sigma X_t^{\rho} dW_t, \tag{7.6}$$

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

TABLE 2. Sizes and powers of the Hong and Li (2005) test under DGPs A0–A5

	Sample size											
	$T = 250$				$T = 500$				$T = 1,000$			
	Significance level				Significance level				Significance level			
	.10		.05		.10		.05		.10		.05	
	Lag p		Lag p		Lag p		Lag p		Lag p		Lag p	
	10	20	10	20	10	20	10	20	10	20	10	20
Sizes												
Vasicek (low persistence)	.155	.146	.104	.103	.140	.136	.079	.094	.137	.151	.082	.095
Vasicek (high persistence)	.128	.153	.087	.102	.145	.145	.104	.095	.127	.132	.082	.092
Powers												
CIR	.170	.152	.116	.106	.288	.276	.206	.208	.576	.550	.456	.440
Ahn and Gao (1999)	.824	.782	.770	.728	.990	.994	.990	.984	1.00	1.00	1.00	1.00
Chan et al. (1992)	.674	.660	.634	.612	.938	.930	.922	.914	1.00	.998	.996	.992
Aït-Sahalia (1996a)	.956	.954	.944	.932	.992	.996	.992	.996	1.00	1.00	1.00	1.00
Jump diffusion	.780	.766	.754	.738	.942	.960	.924	.938	.998	.996	.998	.996

Note: DGP A0 is the Vasicek (1977) model, given in equation (7.1); low persistence and high persistence correspond to $(K, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$ and $(0.214592, 0.089102, 0.000546)$, respectively; DGPs A1–A5 are the CIR model, the Ahn and Gao (1999) inverse-Feller model, the Chan et al. (1992) model, the Aït-Sahalia (1996a) nonlinear drift model, and the jump diffusion model, given in equations (7.4)–(7.8), respectively. Results are based on the Hong and Li (2005) test. The p values of sizes are based on the results of 1,000 iterations using asymptotic critical values; the p values of powers are based on the results of 500 iterations using empirical critical values.

DGP A4 (Aït-Sahalia, 1996a, nonlinear drift model).

$$dX_t = \left(\alpha_{-1} X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2\right) dt + \sigma X_t^{\rho} dW_t, \tag{7.7}$$

where $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma^2, \rho) = (0.00107, -0.0517, 0.877, -4.604, 0.64754, 1.50)$.[19]

DGP A5 (AJD model).

$$dX_t = \kappa(\alpha - X_t)\, dt + \sigma\, dW_t + J_t dN_t, \tag{7.8}$$

where $(\kappa, \alpha, \sigma^2) = (0.214592, 0.089102, 0.001986)$, N_t is a Poisson process with the intensity $\lambda = 0.05$, and J_t is the jump size, which is independent of N_t and has a normal distribution $N(0, \eta^2 = 0.003973)$.

As in 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 are taken from Ahn and Gao (1999). For DGPs A3 and A4, the parameter values are taken from the Aït-Sahalia (1999) estimates of real interest rate data. For the jump diffusion model, the parameter values are calculated from model (7.1) using the Aït-Sahalia et al. (2009) method in their Example 3. For each of these four alternatives, we generate 500 data sets of the random sample $\{X_t\}_{t=\Delta}^{T\Delta}$, where $T = 250$, 500, and 1,000, respectively, at the monthly sample frequency. For the CIR, Ahn and Gao, and AJD models, we simulate data from model transition densities, which have closed forms. For the Chan et al. (1992) and Aït-Sahalia (1996a) nonlinear drift models, whose transition densities have no closed form, we simulate data by the Euler scheme. Each simulated sample path is generated using 120 intervals per month. We then discard 119 out of every 120 observations, obtaining discrete observations at the monthly frequency.

For each data set, we use MLE to estimate model (7.1). Table 3 reports the rejection rates of $\hat{M}$ and $\hat{M}^{(\nu)}$ at the 10% and 5% levels using empirical critical values, which are obtained under $\mathbb{H}_0$ and provide fair comparison on an equal ground. Again, we include tests using bootstrap critical values when $T = 250$. Under DGP A1, model (7.1) is correctly specified for the drift function but is misspecified for the diffusion function because it fails to capture the "level effect." Both asymptotic and finite-sample versions of the omnibus test have reasonable power under DGP A1, with rejection rates around 70% at the 5% level when $T = 1{,}000$. The finite-sample $\hat{M}_{FS}$ has slightly higher rejection rates than the asymptotic version of the $\hat{M}$ test. The Hong and Li (2005) test is less powerful than the $\hat{M}$ tests, with rejection rates around 45% at the 5% level when $T = 1{,}000$. The variance test $\hat{M}^{(2)}$ has good power, and the rejection rates increase with T. Interestingly, the mean test $\hat{M}^{(1)}$ has no power, indicating that these diagnostic tests do not overreject the correctly specified conditional mean dynamics.

Under DGP A2, model (7.1) is misspecified for both the conditional mean and conditional variance because it ignores the nonlinear drift and diffusion. As expected, both asymptotic and finite-sample versions of the omnibus $\hat{M}$ test have good power when model (7.1) is used to fit the data generated from DGP A2. The power of $\hat{M}$ increases significantly with T and approaches unity when $T = 1{,}000$. The Hong and Li (2005) test is more powerful than the $\hat{M}$ tests in small samples, but the difference becomes smaller as T increases. Both the mean test $\hat{M}^{(1)}$ and the variance test $\hat{M}^{(2)}$ have power, and the rejection rates increase with T.

Under DGP A3, the diffusion is no longer a linear function of X_t. Thus model (7.1) is misspecified for the conditional mean and conditional variance. Both the asymptotic and finite-sample versions of the omnibus $\hat{M}$ test have good power when model (7.1) is used to fit the data generated from DGP A3. The rejection rates approach unity when $T = 1{,}000$. However, the mean test $\hat{M}^{(1)}$ has little power in detecting mean misspecification. One conjecture is that the difference between the true conditional mean under DGP A3 and the model (7.1)–implied conditional mean is small. This can be seen from a discretized version of DGP A3, namely, $X_{t+\Delta} \approx X_t + \left[\kappa(\alpha - X_t) - \frac{1}{2}\sigma^2 X_t^{2\rho}\right]\triangle + \sigma X_t^{\rho} \triangle W_t + \frac{1}{2}\sigma^2 X_t^{2\rho}(\triangle W_t)^2$,

TABLE 3. Powers of specification tests under DGPs A1–A5

	Significance level							
	.10		.05		.10		.05	
	BCV-AS	BCV-FS	BCV-AS	BCV-FS	ECV-AS	ECV-FS	ECV-AS	ECV-FS
	$T = 250$				$T = 250$			
CIR								
$\hat{M}$	.428	.454	.270	.266	.122	.126	.048	.058
$\hat{M}^{(1)}$	.134	.134	.076	.070	.016	.018	.004	.006
$\hat{M}^{(2)}$	.326	.314	.180	.132	.106	.084	.058	.044
Ahn and Gao (1999)								
$\hat{M}$	.848	.888	.688	.740	.548	.594	.378	.408
$\hat{M}^{(1)}$	.306	.336	.202	.218	.078	.070	.032	.034
$\hat{M}^{(2)}$	.544	.438	.424	.344	.294	.218	.216	.154
Chan et al. (1992)								
$\hat{M}$	.506	.554	.324	.370	.204	.248	.108	.134
$\hat{M}^{(1)}$	.196	.214	.110	.120	.034	.040	.014	.020
$\hat{M}^{(2)}$	.356	.294	.240	.200	.174	.132	.104	.076
Aït-Sahalia (1996a)								
$\hat{M}$	.818	.906	.612	.754	.632	.732	.482	.578
$\hat{M}^{(1)}$	.222	.316	.150	.184	.176	.214	.112	.128
$\hat{M}^{(2)}$	.318	.372	.216	.260	.286	.318	.192	.212
Jump diffusion								
$\hat{M}$	.580	.590	.538	.564	.604	.622	.536	.592
$\hat{M}^{(1)}$	.130	.068	.076	.042	.116	.074	.084	.048
$\hat{M}^{(2)}$	.374	.076	.274	.044	.372	.100	.274	.046
	ECV-AS	ECV-FS	ECV-AS	ECV-FS	ECV-AS	ECV-FS	ECV-AS	ECV-FS
	$T = 500$				$T = 1{,}000$			
CIR								
$\hat{M}$	.354	.368	.206	.214	.860	.890	.710	.726
$\hat{M}^{(1)}$	.012	.014	.006	.008	.026	.028	.004	.004
$\hat{M}^{(2)}$	.278	.240	.162	.118	.440	.406	.312	.268
Ahn and Gao (1999)								
$\hat{M}$	.974	.980	.912	.944	1.00	1.00	1.00	1.00
$\hat{M}^{(1)}$	.138	.170	.076	.084	.252	.286	.126	.180
$\hat{M}^{(2)}$	.532	.432	.396	.318	.816	.764	.744	.682
Chan et al. (1992)								
$\hat{M}$	.666	.706	.434	.504	.986	.996	.936	.952
$\hat{M}^{(1)}$	.096	.116	.048	.062	.134	.166	.084	.104
$\hat{M}^{(2)}$	.174	.132	.104	.076	.582	.526	.470	.422
Aït-Sahalia (1996a)								
$\hat{M}$	.922	.952	.788	.870	.998	1.00	.992	.996
$\hat{M}^{(1)}$	.142	.196	.100	.126	.106	.142	.052	.078
$\hat{M}^{(2)}$	.286	.318	.192	.212	.484	.456	.348	.334
Jump diffusion								
$\hat{M}$	.754	.752	.738	.738	.840	.840	.832	.832
$\hat{M}^{(1)}$	.148	.110	.110	.054	.160	.094	.082	.048
$\hat{M}^{(2)}$	.712	.106	.612	.068	.918	.168	.872	.128

Note: $\hat{M}$, $\hat{M}^{(1)}$, and $\hat{M}^{(2)}$ are the omnibus test, the conditional mean test, and the conditional variance test, respectively; BCV-AS and BCV-FS denote the asymptotic version and the finite-sample version using bootstrap critical values, respectively; ECV-AS and ECV-FS denote the asymptotic version and the finite-sample version using empirical critical values, respectively. Number of iterations = 500.

where $\triangle$ is the length of the time discretization subinterval and $\triangle W_t$ is the increment of the Brownian motion. With small X_t (so that $X_t^{2\rho}$ is negligible for $\rho = 1.46$), the leading term that determines the true conditional mean under DGP A3 is $\kappa(\alpha - X_t)$, which coincides with the drift of model (7.1). Thus, our mean test $\hat{M}^{(1)}$ has little power in detecting the small differences between the null and alternative models. The variance test $\hat{M}^{(2)}$, however, has power, and the rejection rates increase with T. The Hong and Li (2005) test is more powerful than the $\hat{M}$ tests in small samples, but the difference becomes smaller as T increases.

Under DGP A4, model (7.1) is misspecified for the conditional mean and conditional variance because it ignores nonlinearity in both drift and diffusion. The patterns of our omnibus and diagnostic tests are similar to those under DGP A3. The Hong and Li (2005) test is more powerful than the $\hat{M}$ tests when $T = 250$, but the rejection rates of the new tests increase quickly with T and approach unity when $T = 1{,}000$.

As shown in Aït-Sahalia et al. (2009), the transition density of DGP A5 is, at the first order in Δ, a mixture of normal distributions: $(1-\lambda\Delta)N\left(\mu_t, \sigma_t^2\right) + \lambda\Delta N\left(\mu_t, \sigma_t^2 + \eta^2\right)$, where μ_t and σ_t^2 are given in (7.2) and (7.3). Under DGP A5, the conditional mean of model (7.1) is correctly specified, but the conditional variance is misspecified. Both the $\hat{M}$ tests and the Hong and Li (2005) test have good power against this jump alternative, and the Hong and Li test is more powerful. We note that the asymptotic version of our variance test $\hat{M}_{AS}^{(2)}$ has good power in detecting variance misspecification but the finite-sample version $\hat{M}_{FS}^{(2)}$ has puzzlingly little power. In most cases, tests using bootstrap critical values have better power than tests using empirical critical values.[20]

7.2. Bivariate Models

7.2.1. Size of the $\hat{M}$ Tests. To examine the size of $\hat{M}$ for bivariate models, we consider the following DGP.

DGP B0 (Bivariate uncorrelated Gaussian diffusion).

$$d\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{22} \end{pmatrix} \begin{pmatrix} \theta_1 - X_{1,t} \\ \theta_2 - X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} d\begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}. \tag{7.9}$$

We set $(\kappa_{11}, \kappa_{22}, \theta_1, \theta_2, \sigma_{11}, \sigma_{22}) = (0.2, 0.8, 0, 0, 1, 1)$.[21] With a diagonal matrix $\boldsymbol{\kappa} = \text{diag}(\kappa_{11}, \kappa_{22})$, DGP B0 is an uncorrelated two-factor Gaussian diffusion process. As shown in Duffee (2002), the Gaussian diffusion model has analytic expressions for the conditional mean and conditional variance, respectively:

$$\mathrm{E}(\mathbf{X}_{t+\Delta}|\mathbf{X}_t) = \left(\mathbf{I} - e^{-\boldsymbol{\kappa}\Delta}\right)\boldsymbol{\theta} + e^{-\boldsymbol{\kappa}\mathbf{1}}\mathbf{X}_t, \tag{7.10}$$

$$\mathrm{var}(\mathbf{X}_{t+\Delta}|\mathbf{X}_t) = \text{diag}\left[\frac{\sigma_{11}^2}{2\kappa_{11}}\left(1 - e^{-2\kappa_{11}\Delta}\right), \frac{\sigma_{22}^2}{2\kappa_{22}}\left(1 - e^{-2\kappa_{22}\Delta}\right)\right], \tag{7.11}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)'$. We simulate 1,000 data sets of the random sample $\{\mathbf{X}_t\}_{t=\Delta}^{T\Delta}$ at the monthly frequency for $T = 250$, 500, 1,000, and 2,500, respectively, from a bivariate normal distribution. For each data set, we use MLE to estimate model (7.9), with no restrictions on the intercepts, and compute the $\hat{M}$ and Hong and Li (2005) statistics.

We focus on the finite-sample version of the omnibus test in the bivariate case, which gives better sizes in finite samples than the asymptotic version of the omnibus test. To reduce computational costs, we generate $\tilde{\mathbf{u}}$ from an $N(\mathbf{0}, \mathbf{I}_2)$ distribution, with each $\tilde{\mathbf{u}}$ having 15 grid points in $\mathbb{R}^2$, and let $\mathbf{u} = (\tilde{\mathbf{u}}', -\tilde{\mathbf{u}}')'$ to ensure its symmetry. We standardize each component of $\mathbf{X}_t$ and choose $h = T^{-1/6}$, which attains the optimal rate for bivariate local linear regression fitting. The calculation of PITs for the bivariate model (7.9) used in Hong and Li (2005) is described in Section 2 (see also (18) and (19) of Hong and Li 2005).

Table 4 reports the rejection rates of $\hat{M}$, $\hat{M}^{(\boldsymbol{\nu})}$, and the Hong and Li (2005) test under DGP B0 at the 10% and 5% levels, using asymptotic theory. The $\hat{M}$ test tends to underreject a bit, and the Hong and Li test tends to overreject a bit. With $|\boldsymbol{\nu}| = 1, 2$, the diagnostic tests $\hat{M}^{(\boldsymbol{\nu})}$ check model misspecifications in conditional means, conditional variances, and conditional correlation of state variables. The $\hat{M}^{(\boldsymbol{\nu})}$ tests tend to overreject a bit but not excessively. Speaking overall, both the omnibus and diagnostic tests have reasonable sizes at the 10% and 5% levels for sample sizes as small as $T = 250$ (i.e., about 20 years of monthly data). Our results show that the reasonable size performance of $\hat{M}$ and $\hat{M}^{(\boldsymbol{\nu})}$ in the univariate models carries over to the bivariate models. We also consider tests using bootstrap critical values for $T = 250$, which provide better sizes than asymptotic theory.

7.2.2. Power of the $\hat{M}$ Tests. To investigate the power of $\hat{M}$ and $\hat{M}^{(\boldsymbol{\nu})}$ in distinguishing model (7.9) from alternative models, we also generate data from four bivariate affine diffusion models, respectively:

DGP B1 (Bivariate correlated Gaussian diffusion, with constant correlation in diffusion).

$$d\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.8 \end{pmatrix} \begin{pmatrix} -X_{1,t} \\ -X_{2,t} \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0.8 & 1 \end{pmatrix} d\begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}. \qquad \textbf{(7.12)}$$

DGP B2 (Bivariate correlated Gaussian diffusion, with constant correlation in drift).

$$d\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0.8 & 0.8 \end{pmatrix} \begin{pmatrix} -X_{1,t} \\ -X_{2,t} \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} d\begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}. \qquad \textbf{(7.13)}$$

DGP B3 (Bivariate Dai and Singleton, 2000, $A_1(2)$ process).

$$d\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.8 \end{pmatrix} \begin{pmatrix} 2 - X_{1,t} \\ -X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \sqrt{X_{1,t}} & 0 \\ 0 & 1 \end{pmatrix} d\begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}. \qquad \textbf{(7.14)}$$

TABLE 4. Sizes of specification tests under DGP B0

		$T=250$				$T=500$		$T=1,000$		$T=2,500$	
	α	.10		.05		.10	.05	.10	.05	.10	.05
		BCV-FS	ACV-FS	BCV-FS	ACV-FS	ACV-FS	ACV-FS	ACV-FS	ACV-FS	ACV-FS	ACV-FS
	$\hat{M}$	.132	.098	.076	.051	.075	.046	.064	.043	.079	.046
$\hat{M}^1$	$\hat{M}^{(1,0)}$	.110	.172	.056	.113	.083	.051	.125	.077	.107	.063
	$\hat{M}^{(0,1)}$	.146	.169	.090	.118	.075	.046	.131	.085	.111	.077
$\hat{M}^2$	$\hat{M}^{(2,0)}$	.118	.095	.052	.053	.084	.060	.130	.089	.119	.079
	$\hat{M}^{(2,0)}$	.130	.090	.072	.056	.084	.047	.102	.071	.114	.073
$\hat{M}^3$	$\hat{M}^{(1,1)}$	.098	.107	.052	.066	.105	.068	.129	.093	.129	.083
						ACV					
Hang and Li (2005)		.167		.105		.176	.110	.144	.092	.146	.106

Note: DGP B0 is a bivariate uncorrelated Gaussian diffusion process, given in equation (7.9); $\hat{M}$ is the omnibus test; $\hat{M}^1, \hat{M}^2, \hat{M}^3$ are conditional mean tests, conditional variance tests, and conditional correlation test, respectively; BCV-FS and ACV-FS denote the finite-sample version tests using bootstrap and asymptotic critical values respectively. The p values of BCV-FS are based on the results of 500 iterations; the p values of ACV-FS and Hong and Li (2005) are based on the results of 1,000 iterations.

DGP B4 (Bivariate correlated diffusion, with time-varying correlation in diffusion).

$$d\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.8 \end{pmatrix}\begin{pmatrix} 2-X_{1,t} \\ 1-X_{2,t} \end{pmatrix}dt + \begin{pmatrix} 1 & 0 \\ 0.5\sqrt{X_{1,t}} & 1 \end{pmatrix} d\begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}. \qquad \textbf{(7.15)}$$

We use the Euler scheme to simulate 500 data sets of the random sample $\{\mathbf{X}_t\}_{t=\Delta}^{T\Delta}$ at the monthly frequency for $T = 250$, 500, and 1,000, respectively. For each data set, we use MLE to estimate model (7.9), with no restrictions on intercept coefficients. Table 5 reports the rejection rates of $\hat{M}_{FS}$, $\hat{M}_{FS}^{(v)}$, and the Hong and Li (2005) test at the 10% and 5% levels using empirical critical values. The empirical critical values are obtained under DGP B0.

With a nondiagonal matrix **6**, DGP B1 is a bivariate correlated Gaussian diffusion process with constant correlation in diffusion. Under DGP B1, model (7.9) ignores the nonzero constant correlation between state variables. The $\hat{M}_{FS}$ test has good power in detecting misspecification in the joint dynamics, with its rejection rate around 43% at the 5% level when $T = 1{,}000$. Interestingly, the Hong and Li (2005) test has no power with rejection rates around significance levels. This is not surprising because the conditional densities of individual variables $p(X_{1,t}, t|\mathcal{I}_{t-1}, X_{2,t}, \boldsymbol{\theta})$ and $p(X_{2,t}, t|\mathcal{I}_{t-1}, \boldsymbol{\theta})$ are correctly specified although the joint dynamics is misspecified. Our correlation test $\hat{M}_{FS}^{(1,1)}$ has good power against correlation misspecification.

DGP B2 is another bivariate correlated Gaussian diffusion process, where the correlation between state variables comes from drifts rather than diffusions. Under DGP B2, model (7.9) is correctly specified for the diffusion function but is misspecified for the drift function. The power patterns of the $\hat{M}_{FS}$ and Hong and Li (2005) tests against the bivariate Vasicek model (7.9) are very similar to those under DGP B1. The rejection rate of $\hat{M}_{FS}$ increases with T and approaches unity when $T = 1{,}000$, whereas the power of the Hong and Li test is close to 5% at the 5% level. The conditional mean and variance of $X_{2,t}$ and the conditional correlation between $X_{1,t}$ and $X_{2,t}$ are misspecified, and our diagnostic tests are able to detect them.

DGP B3 is the Dai and Singleton (2000) $A_1(2)$ model, where the first factor affects the instantaneous variance of $\mathbf{X}_t$. Under DGP B3, model (7.9) is correctly specified for the drift function but is misspecified for the diffusion function because it fails to capture the "level effect" of $X_{1,t}$. Both the $\hat{M}_{FS}$ and Hong and Li (2005) tests have excellent power under DGP B3. In small samples, $\hat{M}_{FS}$ is more powerful than the Hong and Li test, but their rejection rates become very close when $T = 1{,}000$. The variance test $\hat{M}_{FS}^{(2,0)}$ has power against the misspecification in conditional variance of $X_{1,t}$. The mean test $\hat{M}_{FS}^{(1,0)}$ tends to overreject a bit although the conditional mean of $X_{1,t}$ is correctly specified. Nevertheless, the overrejection is not severe. The mean test $\hat{M}_{FS}^{(0,1)}$, the variance test $\hat{M}_{FS}^{(0,2)}$, and the correlation test $\hat{M}_{FS}^{(1,1)}$ do not overreject correctly specified conditional moments.

TABLE 5. Powers of specification tests under DGPs B1–B4

		Sample size							
		$T = 250$				$T = 500$		$T = 1,000$	
		Significance level				Significance level		Significance level	
		.10		.05		.10	.05	.10	.05
		BCV-FS	ECV-FS	BCV-FS	ECV-FS	ECV-FS	ECV-FS	ECV-FS	ECV-FS
DGP B1 (Bivariate correlated Gaussian diffusion process, with constant correlation in diffusion)									
	$\hat{M}$	.278	.210	.184	.124	.285	.166	.601	.435
Hong and Li (2005)			.080		.052	.112	.066	.132	.070
$\hat{M}^1$	$\hat{M}^{(1,0)}$	.110	.098	.054	.054	.108	.064	.118	.068
	$\hat{M}^{(0,1)}$	.100	.082	.042	.038	.084	.048	.094	.041
$\hat{M}^2$	$\hat{M}^{(2,0)}$	.096	.086	.052	.048	.116	.058	.108	.044
	$\hat{M}^{(2,0)}$	.086	.108	.046	.056	.068	.036	.104	.060
$\hat{M}^3$	$\hat{M}^{(1,1)}$	.516	.515	.368	.363	.641	.517	.862	.748
DGP B2 (Bivariate correlated Gaussian diffusion process, with constant correlation in drift)									
	$\hat{M}$	.454	.489	.342	.311	.898	.846	1.00	.998
Hong and Li (2005)			.074		.046	.078	.040	.100	.066
$\hat{M}^1$	$\hat{M}^{(1,0)}$	.080	.054	.048	.032	.108	.058	.116	.052
	$\hat{M}^{(0,1)}$	.660	.637	.544	.481	.952	.930	1.00	.998
$\hat{M}^2$	$\hat{M}^{(2,0)}$	.118	.122	.066	.062	.114	.050	.050	.020
	$\hat{M}^{(2,0)}$	.362	.391	.242	.255	.759	.625	.962	.904
$\hat{M}^3$	$\hat{M}^{(1,1)}$	.338	.409	.214	.269	.778	.619	.978	.926
DGP B3 (Dai and Singleton (2000) $A_1(2)$ process)									
	$\hat{M}$	.720	.816	.618	.754	.946	.900	.998	.990
Hong and Li (2005)			.728		.688	.946	.936	1.00	1.00
$\hat{M}^1$	$\hat{M}^{(1,0)}$	.166	.216	.102	.136	.164	.096	.196	.110
	$\hat{M}^{(0,1)}$	.122	.114	.056	.050	.114	.064	.092	.054
$\hat{M}^2$	$\hat{M}^{(2,0)}$	.268	.252	.184	.190	.373	.261	.459	.285
	$\hat{M}^{(2,0)}$	.102	.102	.068	.050	.130	.062	.108	.046
$\hat{M}^3$	$\hat{M}^{(1,1)}$	.110	.098	.044	.058	.092	.046	.064	.028
DGP B4 (Bivariate correlated diffusion process, with time-varying correlation in diffusion)									
	$\hat{M}$	.106	.214	.060	.144	.590	.446	.982	.958
Hong and Li (2005)			.130		.094	.166	.110	.122	.076
$\hat{M}^1$	$\hat{M}^{(1,0)}$	.114	.060	.054	.034	.196	.132	.254	.162
	$\hat{M}^{(0,1)}$	.126	.044	.068	.012	.106	.054	.142	.084
$\hat{M}^2$	$\hat{M}^{(2,0)}$	.138	.134	.076	.070	.230	.158	.286	.190
	$\hat{M}^{(2,0)}$	.126	.098	.052	.052	.114	.056	.152	.080
$\hat{M}^3$	$\hat{M}^{(1,1)}$	.454	.474	.328	.344	.724	.594	.912	.818

Note: DGP B1 is a bivariate correlated Gaussian diffusion model, with constant correlation in diffusion, given in equation (7.12). DGP B2 is a bivariate correlated Gaussian diffusion model, with constant correlation in drift, given in equation (7.13). DGP B3 is the Dai and Singleton (2000) $A_1(2)$, given in equation (7.14). DGP B4 is a bivariate correlated diffusion model, with time-varying correlation in diffusion, given in equation (7.15). The p values are based on the results of 500 iterations. BCV-FS and ECV-FS denote the finite-sample version tests using bootstrap and empirical critical values, respectively.

DGP B4 is a bivariate time-varying correlated Gaussian diffusion process, where the correlation depends on $X_{1,t}$. If we use model (7.9) to fit data generated from DGP B4, there exists dynamic misspecification in conditional covariance between state variables. The $\hat{M}_{FS}$ test has good power when (7.9) is used to fit data generated from DGP B4. The rejection rate of the $\hat{M}_{FS}$ test increases to 95.8% at the 5% level when $T = 1{,}000$. The power of the Hong and Li (2005) test is still close to 5% at the 5% level. The correlation test $\hat{M}_{FS}^{(1,1)}$ has good power against this dynamic correlation misspecification, as the rejection rate is about 82% at the 5% level when $T = 1{,}000$.

To sum up, we make the following observations: (1) For both univariate and bivariate models, the $\hat{M}$ and $\hat{M}^{(\nu)}$ tests have reasonable sizes in finite samples, particularly when the parametric bootstrap is used. The finite-sample versions of the proposed tests have better sizes than the asymptotic versions of the tests. (2) The omnibus test $\hat{M}$ has reasonable omnibus power in detecting various model missspecifications. It has reasonable power even when the sample size T is as small as 250. It has advantages in a multivariate framework. Particularly, it has good power in detecting misspecification in the joint dynamics even when the dynamics of individual components is correctly specified. This feature is not attainable by the Hong and Li (2005) test. (The Hong and Li test performs well in the univariate case.) (3) The directional diagnostic tests $\hat{M}^{(\boldsymbol{\nu})}$ can check various specific aspects of model misspecifications. Generally speaking, the mean test $\hat{M}^{(\boldsymbol{\nu})}$, with $|\boldsymbol{\nu}| = 1$, can detect misspecification in drifts; the variance test $\hat{M}^{(\boldsymbol{\nu})}$, with $|\boldsymbol{\nu}| = 2$, can check misspecifications in variances and correlations, respectively. However, the mean test may fail to detect mean misspecification if the discrepancy between the data-implied conditional mean and the model-implied conditional mean is small.

8. CONCLUSION

The CCF-based estimation of continuous-time multivariate Markov models has attracted increasing attention in financial econometrics. We have complemented this literature by proposing a CCF-based nonparametric regression omnibus test for the adequacy of a continuous-time multivariate Markov model. Our omnibus test fully exploits the information in the joint dynamics of state variables and thus can capture misspecification in modeling the joint dynamics, which may be easily missed by existing procedures. In addition, our omnibus test exploits the Markov property under both the null and alternative hypotheses and is an efficient testing approach when the DGP is indeed Markov. A class of diagnostic procedures is supplemented to gauge possible sources of model misspecifications. All test statistics follow an asymptotic null $N(0,1)$ distribution and allow for data-driven bandwidth sequences, and they are applicable to various estimation methods, including suboptimal but consistent estimators. Simulation studies show that the proposed tests perform reasonably in finite samples for both univariate and bivariate continuous-time models.

NOTES

1. This is the basic idea behind Markov decision processes (MDPs), which provide a broad framework for modeling sequential decision making under uncertainty. MDPs have been used extensively in both microeconomics and macroeconomics and also in finance and marketing (for an excellent survey, see, e.g., Rust, 1994). Applications include investment under uncertainty (Lucas and Prescott, 1971), asset pricing models (Hall, 1978), economic growth (Lucas, 1988), optimal taxation (Lucas and Stokey, 1983), and equilibrium business cycles (Kydland and Prescott, 1982).

2. In a review, Sundaresan (2001) states 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."

3. Aït-Sahalia (1996a) also proposes a transition density–based test that exploits the "transition discrepancy" characterized by the forward and backward Kolmogorov equations, although the marginal density–based test is more emphasized there.

4. Modeling the joint dynamics, especially correlations, has become increasingly important in many financial applications such as pricing and risk management (e.g., Dai and Singleton, 2000). As Engle (2002) points out, "the quest for reliable estimates of correlations between financial variables has been the motivation for countless academic articles, practitioner conferences and Wall Street research."

5. It is assumed that μ, σ, ν, and λ are regular enough to have a unique strong solution to (2.1). See, e.g., Aït-Sahalia (1996a) and Duffie et al. (2000).

6. We use the symmetry of $W(\cdot)$ to simplify the expression of the asymptotic variance of $L^2(\hat{m})$.

7. Parameter estimation uncertainty may have an impact on finite-sample performance when $\{\mathbf{X}_t\}$ is highly persistent. A parametric bootstrap can be used to capture this impact. See the simulation studies in Section 7.

8. The Bahadur relative efficiency is defined as the limiting ratio of the sample sizes required by the two tests under comparison to achieve the same asymptotic significance level (P-value) under the same fixed alternative.

9. Assuming that J_t in (2.1) is a Poisson process, Yu (2007) shows that $\Pr(A_{t,\Delta}|\mathbf{X}_{t-\Delta}, \boldsymbol{\theta}) = O(\Delta)$ and $\Pr(A^c_{t,\Delta}|\mathbf{X}_{t-\Delta}, \boldsymbol{\theta}) = O(1)$, where $A_{t,\Delta}$ denotes the event of jump that occurs between time $t-\Delta$ and t and $A^c_{t,\Delta}$ denotes its complement.

10. We can also define $\phi(\mathbf{u}_1, t|\mathbf{X}_{1,t-1}, \boldsymbol{\theta}) \equiv \mathrm{E}_{\boldsymbol{\theta}}\left[\exp\left(i\mathbf{u}_1'\mathbf{X}_{1,t}\right)|\mathbf{X}_{1,t-1}\right] = \int \varphi[(\mathbf{u}_1', \mathbf{0}')', t|\mathbf{X}_{1,t-1}, \mathbf{x}_{2,t-1}, \boldsymbol{\theta}] p\left(\mathbf{x}_{2,t-1}, \boldsymbol{\theta}\right) d\mathbf{x}_{2,t-1}$ and $\varepsilon_{1,t} = \exp\left(i\mathbf{u}_1'\mathbf{X}_{1,t}\right) - \phi(\mathbf{u}_1, t|\mathbf{X}_{1,t-1}, \boldsymbol{\theta})$. Then under $\mathbb{H}_0$, we have $\mathrm{E}\left[\varepsilon_{1,t}(\mathbf{u}_1, \boldsymbol{\theta}_0)|\mathbf{X}_{t-1}\right] = 0$ a.s. for all $\mathbf{u}_1 \in \mathbb{R}^{d_1}$ and some $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. A test can be constructed correspondingly. This characterization has been used in Chacko and Viceira (2003) for estimation. As Chacko and Viceira (2003) point out, this approach does not condition on the entire path of the observable but only conditions on the observable in the previous period. "By not conditioning on this information we lose efficiency, but the trade-off is that we gain immensely in terms of computational speed." Similarly, the estimation of $\hat{\phi}(\mathbf{u}_1, t|\mathbf{X}_{1,t-1}, \boldsymbol{\theta})$ is much easier, as we can simulate $\mathbf{X}_{2,t}$ from the marginal density $p\left(\mathbf{x}_{2,t}, \boldsymbol{\theta}\right)$ and use the sample average to approximate the integral.

11. Chen and Hong (2005) can be applied here because $Z_{1,t}(\mathbf{u}_1, \boldsymbol{\theta}_0)$ is generally not Markov. But we shall propose an alternative test, using a nonparametric regression approach.

12. Based on their methodology, Christoffersen, Jacobs, and Minouni (2009) estimate models with particle filters for option pricing application, and Golightly and Wilkinson (2006) use time discretizations and particle filters for estimating diffusion models.

13. This is called sampling/importance resampling (SIR) in the literature. Alternative methods include rejection sampling and the MCMC althorithm. See Doucet, de Freitas, and Gordon (2001) and Pitt and Shephard (1999) for more discussion.

14. Examples include AJD models (Duffie et al., 2000) and time-changed Lévy processes (Carr and Wu, 2003, 2004).

15. We also simulate data sets at the weekly and daily frequencies ($\Delta = \frac{1}{52}, \frac{1}{250}$, respectively), and simulation patterns are similar. These results are available from the authors upon request.

16. The bootstrap procedure is valid even if we do not reestimate the model in step (ii). But the current procedure is expected to yield better finite-sample performance as it mimics the actual distribution of the nonbootstrap test statistic better under the null hypothesis by taking into account parameter estimation uncertainty. A bit surprisingly, our simulation experiment shows that the sizes are close although the current procedure yields better power. We conjecture that the impact of estimation uncertainty is small but the degenerate U-statistic is not approximated well by the normal distribution.

17. The proof is available from the authors upon request.

18. There are some typos in the parameter values of the Ahn and Gao (1999) inverse-Feller model used in the Hong and Li (2005) simulation study. We have corrected them correspondingly.

19. We note that DGPs A2–A4 do not satisfy the conditions in Proposition 9 in Hansen and Scheinkman (1995), which provides the sufficient condition for a mixing process. As the mixing condition is a sufficient but not necessary condition for our asymptotic theory, we verify the validity of our tests via simulation.

20. We conjecture that it is due to the different dynamic structures that are used to generate empirical critical values and bootstrap critical values. Empirical critical values are obtained under DGPs A0 and B0, respectively, with 1,000 iterations. Because we estimate the null model under the alternative DGPs, the parameter estimated values may be different from the parameter values under the null DGP A0 or B0. Therefore, bootstrap critical values and empirical critical values may be different.

21. We also try different parameters controlling the degree of persistence. Simulation results show that our tests are not very sensitive to persistence of serial dependence in observations for the bivariate case also.

REFERENCES

Ahn, D., R. Dittmar, & A.R. Gallant (2002) Quadratic term structure models: Theory and evidence. *Review of Financial Studies* 15, 243–288.

Ahn, D. & B. Gao (1999) A parametric nonlinear model of term structure dynamics. *Review of Financial Studies* 12, 721–762.

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

Aït-Sahalia, Y. (1996b) Nonparametric pricing of interest rate derivative securities. *Econometrica* 64, 527–560.

Aït-Sahalia, Y. (1999) Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54, 1361–1395.

Aït-Sahalia, Y. (2002) Maximum-likelihood estimation of discretely sampled diffusions: A closed-form approach. *Econometrica* 70, 223–262.

Aït-Sahalia, P.J. Bickel, & T.M. Stoker (2001) Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics* 105, 363–412.

Aït-Sahalia, Y., J. Fan, & H. Peng (2009) Nonparametric transition-based tests for diffusions. *Journal of the American Statistical Association*, forthcoming.

Andersen, T.G. & J. Lund (1996) Stochastic Volatility and Mean Drift in the Short Rate Diffusion: Sources of Steepness, Level and Curvature in the Yield Curve. Working paper, Northwestern University.

Andrews, D.W.K. (1997) A conditional Kolmogorov test. *Econometrica* 65, 1097–1128.

Bahadur, R.R. (1960) Stochastic comparison of tests. *Annals of Mathematical Statistics* 31, 276–295.

Bally, V. & D. Talay (1996) The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. *Probability Theory and Related Fields* 104, 43–60.

Bates, D. (1996) Jumps and stochastic volatility: Exchange rate processes implicit in PHLX Deutschemake options. *Review of Financial Studies* 9, 69–107.

Bates, D. (2007) Maximum likelihood estimation of latent affine processes. *Review of Financial Studies* 19, 909–965.

Bhardwaj, G., V. Corradi, & N.R. Swanson (2008) A simulation based specification test for diffusion processes. *Journal of Business & Economic Statistics* 26, 176–193.

Brandt, M. & 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.

Carr, P. & L. Wu (2003) The finite moment log stable process and option pricing. *Journal of Finance* 58, 753–777.

Carr, P. & L. Wu (2004) Time-changed Lévy processes and option pricing. *Journal of Financial Economics* 71, 113–141.

Carrasco, M., M. Chernov, J.P. Florens, & E. Ghysels (2007) Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics,* 140, 529–573.

Chacko, G. & L. Viceira (2003) Spectral GMM estimation of continuous-time processes. *Journal of Econometrics* 116, 259–292.

Chan, K.C., G.A. Karolyi, F.A. Longstaff, & A.B. Sanders (1992) An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 47, 1209–1227.

Chen, B. & Y. Hong (2005) Diagnosing Multivariate Continuous-Time Models with Application to Affine Term Structure Models. Working paper, Cornell University.

Chen, S.X., J. Gao, & C.Y. Tang (2008) A test for model specification of diffusion processes. *Annals of Statistics* 36, 162–198.

Chernov, M., A.R. Gallant, E. Ghysels, & G. Tauchen (1999) A New Class of Stochastic Volatility Models with Jumps: Theory and Estimation. Working paper, Columbia University, University of North Carolina at Chapel Hill, and Duke University.

Chib, S., M.K. Pitt, & N. Shephard (2004) Likelihood Based Inference for Diffusion Driven Models. Working paper, University of Oxford.

Christoffersen, P., K. Jacobs, & K. Mimouni (2009) Models for S&P 500 Dynamics: Evidence from Realized Volatility, Daily Returns, and Option Prices. Working paper, McGill University.

Cleveland, W.S. (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74, 829–836.

Cox, J.C., J.E. Ingersoll, & S.A. Ross (1985) A theory of the term structure of interest rates. *Econometrica* 53, 385–407.

Dai, Q. & K. Singleton (2000) Specification analysis of affine term structure models. *Journal of Finance* 55, 1943–1978.

Del Moral, P. & J. Jacod (2001) Interacting particle filtering with discrete observations. In A. Doucet, N. de Freitas, & N. Gordon (eds.), *Sequential Monte Carlo Methods in Practice*, pp. 43–77. Springer.

Del Moral, P., J. Jacod, & P. Protter (2001) The Monte-Carlo method for filtering with discrete-time observations. *Probability Theory and Related Fields* 120, 346–368.

Doucet, A., N. de Freitas, & N. Gordon, eds. (2001) *Sequential Monte Carlo Medhods in Practice*. Springer.

Duffee, G. (2002) Term premia and interest rate forecasts in affine models. *Journal of Finance* 57, 405–443.

Duffie, D. & R. Kan (1996) A yield-factor model of interest rates. *Mathematical Finance* 6, 379–406.

Duffie, D., J. Pan, & K. Singleton (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.

Duffie, D., L. Pedersen, & K. Singleton (2003) Modeling credit spreads on sovereign debt: A case study of Russian bonds. *Journal of Finance* 55, 119–159.

Engle, R. (2002) Dynamic conditional correlation — a simple class of multivariate GARCH models. *Journal of Business & Economic Statistics* 20, 339–350.

Epps, T.W. & L.B. Pulley (1983) A test for normality based on the empirical characteristic function. *Biometrika* 70, 723–726.

Fan, J. (1993) Local linear regression smoothers and their minimax efficiency. *Annals of Statistics* 21, 196–216.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag.

Fan, Y., Q. Li, & I. Min (2006) A nonparametric bootstrap test of conditional distributions. *Econometric Theory* 22, 587–612.

Fenton, V. & A.R. Gallant (1996) Convergence rates of SNP density estimators. *Econometrica* 64, 719–727.

Feuerverger, A. & P. McDunnough (1981) On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society, Series B* 43, 20–27.

Gallant, A.R. & J.R. Long (1997) Estimating stochastic differential equations efficiently by minimum chi-square. *Biometrika* 125–141.

Gallant, A.R. & G. Tauchen (1996) Which moments to match? *Econometric Theory* 12, 657–681.

Gallant, A.R. & G. Tauchen (1998) Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association* 93, 10–24.

Gao, J. & M. King (2004) Adaptive testing in continuous-time diffusion models. *Econometric Theory* 20, 844–882.

Golightly, A. & D. Wilkinson (2006) Bayesian sequential inference for nonlinear multivariate diffusions. *Statistics and Computing* 16, 323–338.

Gordon, N., D. Salmond, & A. Smith (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings* F140, 107–113.

Hall, R. (1978) Stochastic implications of the life cycle–permanent income hypothesis: Theory and practice. *Journal of Political Economy* 86, 971–987.

Hansen, B. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726–748.

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

Hausenblas E. (2002) Error analysis for approximation of stochastic differential equations driven by Poisson random measures. *SIAM Journal of Numerical Analysis* 40, 87–113.

Heston, S. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327–344.

Hjellvik, V., Q. Yao, & D. Tjøhstheim (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.

Hong, Y. & 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.

Jiang, G. & J. Knight (2002) Estimation of continuous-time processes via the empirical characteristic function. *Journal of Business & Economic Statistics* 20, 198–212.

Johannes, M., N. Polson, & J. Stroud (2009) Optimal filtering of jump diffusions: Extracting latent states from asset prices. *Review of Economic Studies*, forthcoming.

Kloeden, P., E. Platen, & H. Schurz (1994) *Numerical Solution of SDE through Computer Experiments*. Springer-Verlag.

Kydland, F.E. & E. Prescott (1982) Time to build and aggregate fluctuations. *Econometrica* 50, 1345–1370.

Li, F. (2007) Testing the parametric specification of the diffusion function in a diffusion process. *Econometric Theory* 23, 221–250.

Li, F. & G. Tkacz (2006) A consistent bootstrap test for conditional density functions with time-series data. *Journal of Econometrics* 133, 841–862.

Lucas, R. (1988) On the mechanics of economic development. *Journal of Monetary Economics* 22, 3–42.

Lucas, R. & E. Prescott (1971) Investment under uncertainty. *Econometrica* 39, 659–681.

Lucas, R. & N.L. Stokey (1983) Optimal fiscal and monetary policy in an economy without capital. *Journal of Monetary Economics* 12, 55–94.

Masry, E. (1996) Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571–599.

Müller, H. (1984) Smooth optimum kernel estimators of densities, regression curves and modes. *Annals of Statistics* 12, 766–774.

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.

Pitt, M.K. & N. Shephard (1999) Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association* 94, 590–599.

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

Rust, J. (1994) Structural estimation of Markov decision processes. In R. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 3081–3143. North-Holland.

Singleton, K. (2001) Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102, 111–141.

Skaug, H.J. & D. Tjøstheim (1996) Testing for serial independence using measures of distance between densities. In P.M. Robinson & M. Rosenblatt (eds.), Athena Conference on Applied Probability and Time Series. Springer Lecture Notes in Statistics 115. Springer.

Stone, C.J. (1977) Consistent nonparametric regression. *Annals of Statistics* 5, 595–645.

Su, L. & H. White (2007) A consistent characteristic function-based test for conditional independence. *Journal of Econometrics* 141, 807–834.

Sundaresan, S. (2001) Continuous-time methods in finance: A review and an assessment. *Journal of Finance* 55, 1569–1622.

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

Yu, J. (2007) Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the chinese yuan. *Journal of Econometrics* 141, 1245–1280.

MATHEMATICAL APPENDIX

We let $\tilde{M}$, $\tilde{C}$, $\tilde{D}$ be defined in the same way as $\hat{M}$, $\hat{C}$, $\hat{D}$ with $\hat{\boldsymbol{\theta}}$ replaced by $\boldsymbol{\theta}_0$ and $\hat{W}_{\hat{h}}$, $\hat{M}_{\hat{h}}$, $\hat{C}_{\hat{h}}$ be defined in the same way as $\hat{W}$, $\hat{M}$, $\hat{C}$ with h replaced by $\hat{h}$. Also, $C \in (1,\infty)$ denotes a generic bounded constant, $||\cdot||$ denotes the usual euclidean norm, and $\frac{\partial}{\partial \boldsymbol{\theta}}\varphi\left(\mathbf{u}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}_0\right)$ denotes $\frac{\partial}{\partial \boldsymbol{\theta}}\varphi\left(\mathbf{u}, t|\mathbf{X}_{t-1}, \boldsymbol{\theta}\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. All convergencies are taken as $T \to \infty$.

Proof of Theorem 1. The proof of Theorem 1 consists of the proofs of Theorems A.1 and A.2.

THEOREM A.1. *Under the conditions of Theorem 1,* $\hat{M} - \tilde{M} \xrightarrow{p} 0$.

THEOREM A.2. *Under the conditions of Theorem 1,* $\tilde{M} \xrightarrow{d} N(0,1)$.

Proof of Theorem A.1. Let $\tilde{m}(\mathbf{u}, \mathbf{X}_{t-1}, \boldsymbol{\theta}_0)$ be defined in the same way as $\hat{m}(\mathbf{u}, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}})$ in (3.3), with $\hat{\boldsymbol{\theta}}$ replaced by $\boldsymbol{\theta}_0$. To show $\hat{M} - \tilde{M} \xrightarrow{p} 0$, it suffices to show

$$\frac{h^{d/2}}{\sqrt{2\hat{D}}} \sum_{t=2}^{T} \int \left[\left| \hat{m}(\mathbf{u}, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}}) \right|^2 - \left| \tilde{m}(\mathbf{u}, \mathbf{X}_{t-1}, \boldsymbol{\theta}_0) \right|^2 \right] a\left(\mathbf{X}_{t-1}\right) dW(\mathbf{u}) = o_P(1), \quad \textbf{(A.1)}$$

$\hat{C} - \tilde{C} = o_P(1)$, and $\hat{D} - \tilde{D} = o_P(1)$. For space, we focus on the proof of (A.1); the proofs for $\hat{C} - \tilde{C} = o_P(1)$ and $\hat{D} - \tilde{D} = o_P(1)$ are tedious but straightforward. We note that it is necessary to obtain the convergence rate $o_P(1)$ for $\hat{C} - \tilde{C}$ to ensure that replacing $\hat{C}$ with $\tilde{C}$ has asymptotically negligible impact.

To show (A.1), we first decompose

$$h^{d/2}\sum_{t=2}^{T}\int\left[\left|\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right|^2-\left|\tilde{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)\right|^2\right]a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})$$
$$=\hat{A}_1+2\mathrm{Re}(\hat{A}_2),\quad \text{where} \tag{A.2}$$
$$\hat{A}_1=h^{d/2}\sum_{t=2}^{T}\int\left|\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})-\tilde{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u}),$$
$$\hat{A}_2=h^{d/2}\sum_{t=2}^{T}\int\left[\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})-\tilde{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)\right]\tilde{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)^* a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u}),$$

where $\mathrm{Re}(\hat{A}_2)$ is the real part of $\hat{A}_2$ and $\tilde{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)^*$ is the complex conjugate of $\tilde{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)$. Then, (A.1) follows from Propositions A.1 and A.2 as $T\to\infty$.

PROPOSITION A.1. *Under the conditions of Theorem 1,* $\hat{A}_1\xrightarrow{p}0$.

PROPOSITION A.2. *Under the conditions of Theorem 1,* $\hat{A}_2\xrightarrow{p}0$.

Proof of Proposition A.1. First we have

$$\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})-\tilde{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)$$
$$=\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\left[\mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)-\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right][1+O_P(\alpha_T)]$$
$$=\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left(\mathbf{u},\mathbf{X}_s|\mathbf{X}_{s-1},\boldsymbol{\theta}_0\right)\right.$$
$$+\frac{1}{2}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\varphi$$
$$\left.\times\left(\mathbf{u},\mathbf{X}_s|\mathbf{X}_{s-1},\bar{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\right][1+O_P(\alpha_T)]$$
$$=\left[\hat{A}_{11}(\mathbf{u},\mathbf{X}_{t-1})+\hat{A}_{12}(\mathbf{u},\mathbf{X}_{t-1})\right][1+O_P(\alpha_T)],\quad \text{say,}$$

where $\bar{\boldsymbol{\theta}}=\lambda\hat{\boldsymbol{\theta}}+(1-\lambda)\boldsymbol{\theta}_0$, $0\le\lambda\le 1$, $\alpha_T=(\ln T/Th)^{1/2}+h$. Note we have used Proposition 6.2 in Fan and Yao (2003, p. 236) and a second-order Taylor series expansion. Proposition A.1 follows from Lemmas A.1 and A.2.

LEMMA A.1. $h^{d/2}\sum_{t=2}^{T}\int\left|\hat{A}_{11}(\mathbf{u},\mathbf{X}_{t-1})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})=o_P(1)$.

LEMMA A.2. $h^{d/2}\sum_{t=2}^{T}\int\left|\hat{A}_{12}(\mathbf{u},\mathbf{X}_{t-1})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})=o_P(1)$.

Proof of Lemma A.1. By Assumptions A.2–A.5, Markov's inequality, and change of variable, we have

$$h^{d/2}\sum_{t=2}^{T}\int\left|\hat{A}_{11}(\mathbf{u},\mathbf{X}_{t-1})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})$$

$$
\begin{aligned}
&=\sum_{t=2}^{T}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^{3}h^{3d/2}g^{2}\left(\mathbf{X}_{t-1}\right)}K(\mathbf{0})^{2}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\varphi(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}_{0})\frac{\partial}{\partial\boldsymbol{\theta}}\varphi^{*}(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}_{0})\right\|dW(\mathbf{u})\\
&\quad\times\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_{0}\right)\right\|^{2}+C\sum_{s>t}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^{3}h^{3d/2}g^{2}\left(\mathbf{X}_{t-1}\right)}K\left(\mathbf{0}\right)K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\\
&\quad\times\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}_{0}\right)\frac{\partial}{\partial\boldsymbol{\theta}}\varphi^{*}\left(\mathbf{u},s|\mathbf{X}_{s-1},\boldsymbol{\theta}_{0}\right)\right\|dW(\mathbf{u})\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_{0}\right)\right\|^{2}\\
&\quad+C\sum_{s>t}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^{3}h^{3d/2}g^{2}\left(\mathbf{X}_{t-1}\right)}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)^{2}\\
&\quad\times\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left(\mathbf{u},s|\mathbf{X}_{s-1},\boldsymbol{\theta}_{0}\right)\right\|^{2}dW(\mathbf{u})\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_{0}\right)\right\|^{2}\\
&\quad+C\sum_{s'>s>t}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^{3}h^{3d/2}g^{2}\left(\mathbf{X}_{t-1}\right)}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)K\left(\frac{\mathbf{X}_{s'-1}-\mathbf{X}_{t-1}}{h}\right)\\
&\quad\times\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left(\mathbf{u},s|\mathbf{X}_{s-1},\boldsymbol{\theta}_{0}\right)\frac{\partial}{\partial\boldsymbol{\theta}}\varphi^{*}\left(\mathbf{u},s|\mathbf{X}_{s'-1},\boldsymbol{\theta}_{0}\right)\right\|dW(\mathbf{u})\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_{0}\right)\right\|^{2}\\
&=O_{P}\left(T^{-2}h^{-3d/2}+T^{-1}h^{-d/2}+T^{-1}h^{-d/2}+h^{d/2}\right)\\
&=o_{P}(1). \qquad \blacksquare
\end{aligned}
$$

Proof of Lemma A.2. By Assumptions A.2–A.5, change of variable, and the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
&h^{d/2}\sum_{t=2}^{T}\int\left|\hat{A}_{12}(\mathbf{u},\mathbf{X}_{t-1})\right|^{2}a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})\\
&\quad\leq\sum_{t=2}^{T}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{2T^{4}h^{3d/2}g^{2}\left(\mathbf{X}_{t-1}\right)}\left|\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\right|^{2}\\
&\quad\times\left[\sup_{\theta\in\boldsymbol{\Theta}}\left\|\frac{\partial^{2}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\varphi\left(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}\right)\right\|\right]^{2}dW(\mathbf{u})\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_{0}\right)\right\|^{4}\\
&\quad\leq C\sum_{t=2}^{T}\frac{a\left(\mathbf{X}_{t-1}\right)K^{2}\left(\mathbf{0}\right)}{T^{4}h^{3d/2}g^{2}\left(\mathbf{X}_{t-1}\right)}+C\sum_{s>t}\frac{a\left(\mathbf{X}_{t-1}\right)K^{2}\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)}{T^{4}h^{3d/2}g^{2}\left(\mathbf{X}_{t-1}\right)}\\
&\quad+C\sum_{s>t}\frac{a\left(\mathbf{X}_{t-1}\right)K\left(0\right)K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)}{T^{4}h^{3d/2}g^{2}\left(\mathbf{X}_{t-1}\right)}
\end{aligned}
$$

$$+C\sum_{s'>s>t}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^4h^{3d/2}g^2\left(\mathbf{X}_{t-1}\right)}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)K\left(\frac{\mathbf{X}_{s'-1}-\mathbf{X}_{t-1}}{h}\right)$$

$$=O_P\left(T^{-3}h^{-3d/2}+T^{-2}h^{-d/2}+T^{-2}h^{-d/2}+T^{-1}h^{d/2}\right)=o_P(1). \qquad \blacksquare$$

Proof of Proposition A.2. We have

$$h^{d/2}\sum_{t=2}^{T}\int\left|\left[\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})-\tilde{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)\right]\tilde{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)^*\right|a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})$$

$$\leq\sum_{t=2}^{T}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^2h^{3d/2}g^2\left(\mathbf{X}_{t-1}\right)}\left|\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\left[\mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)-\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right]\right|$$

$$\times\left|\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\mathbf{Z}_s^*\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|dW(\mathbf{u})\left[1+O_P(\alpha_T)\right]$$

$$=\int\int\frac{a(\mathbf{x})}{h^{d/2}}\left|\hat{A}_{11}(\mathbf{u},\mathbf{x})+\hat{A}_{12}(\mathbf{u},\mathbf{x})\right|$$

$$\times\left|\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right)\mathbf{Z}_s^*\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|d\mathbf{x}\,dW(\mathbf{u})\left[1+O_P(\alpha_T)\right]+o_P(1)$$

$$=O_P(T^{-1/2}h^{-d}+T^{-1}h^{-d/2})+o_P(1)=o_P(1),$$

where we used the fact that $\hat{G}(\mathbf{x})-G(\mathbf{x})=O_P\left(T^{-1/2}(\ln T)^2\right)$ uniformly in $\mathbf{x}\in\mathbf{G}$, and $\hat{G}(\mathbf{x})=T^{-1}\sum_{t=1}^{T}1(\mathbf{X}_t\leq\mathbf{x})$ and $G(\mathbf{x})=P(\mathbf{X}_t\leq\mathbf{x})$. $\blacksquare$

Proof of Theorem A.2. To show $\tilde{M}\xrightarrow{d}N(0,1)$, it suffices to show

$$M\equiv\left[h^{d/2}\sum_{t=2}^{T}\int\left|\frac{1}{Th^dg\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K^*\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|^2\right.$$

$$\left.\times a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})-\tilde{C}\right]\Big/\sqrt{2\tilde{D}}\xrightarrow{d}N(0,1), \qquad \textbf{(A.3)}$$

given Proposition 6.2 in Fan and Yao (2003, p. 236). We shall show Propositions A.3 and A.4 next.

PROPOSITION A.3. *Under the conditions of Theorem 1,*

$$h^{d/2}\sum_{t=2}^{T}\int\left|\frac{1}{Th^dg\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|^2a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})$$

$$=\tilde{C}+2\tilde{U}+o_P(1),$$

where $\tilde{U}=\frac{1}{Th^{3d/2}}\sum_{t=2}^{T}\sum_{t>s}\int\frac{a(\mathbf{x})}{g(\mathbf{x})}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right)K\left(\frac{\mathbf{X}_{t-1}-\mathbf{x}}{h}\right)d\mathbf{x}\int Re[\mathbf{Z}_t(\mathbf{u},\boldsymbol{\theta}_0)\,\mathbf{Z}_s^*(\mathbf{u},\boldsymbol{\theta}_0)]dW(\mathbf{u})$.

PROPOSITION A.4. *Under the conditions of Theorem 1,* $\tilde{U}/\sqrt{\tilde{D}/2} \xrightarrow{d} N(0,1)$.

Proof of Proposition A.3. We first decompose

$$h^{d/2}\sum_{t=2}^{T}\int\left|\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})$$

$$= \frac{1}{T^2h^{3d/2}}\sum_{t=2}^{T}\sum_{s\neq t}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)^2\left|\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|^2 dW(\mathbf{u})$$

$$+\frac{1}{T^2h^{3d/2}}\sum_{t\neq s\neq r}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)K\left(\frac{\mathbf{X}_{r-1}-\mathbf{X}_{t-1}}{h}\right)$$

$$\times\ \mathrm{Re}\left[\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\mathbf{Z}_r^*\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right]dW(\mathbf{u})$$

$$+\frac{1}{T^2h^{3d/2}}\sum_{t=2}^{T}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)}K\left(\mathbf{0}\right)^2\left|\mathbf{Z}_t\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|^2 dW(\mathbf{u})$$

$$+\frac{1}{T^2h^{3d/2}}\sum_{t=2}^{T}\sum_{s\neq t}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)}K\left(\mathbf{0}\right)K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)$$

$$\times\ \mathrm{Re}\left[\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\mathbf{Z}_t^*\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right]dW(\mathbf{u})$$

$$\equiv C_1+U_1+R_1+R_2. \tag{A.4}$$

Next, putting $\boldsymbol{\xi}_t=\left(\mathbf{X}_t,\mathbf{X}_{t-1}\right)$, we define

$$\Phi\left(\boldsymbol{\xi}_t,\boldsymbol{\xi}_s\right)=\int\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)^2\left|\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|^2 dW\left(\mathbf{u}\right)$$

$$+\int\frac{a\left(\mathbf{X}_{s-1}\right)}{g^2\left(\mathbf{X}_{s-1}\right)}K\left(\frac{\mathbf{X}_{t-1}-\mathbf{X}_{s-1}}{h}\right)^2\left|\mathbf{Z}_t\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|^2 dW(\mathbf{u}),$$

and $\Phi\left(\boldsymbol{\xi}_t\right)=\int\Phi\left(\boldsymbol{\xi}_t,\boldsymbol{\xi}_s\right)dP\left(\boldsymbol{\xi}_s\right)$, $\Phi=\int\Phi\left(\boldsymbol{\xi}_t\right)dP\left(\boldsymbol{\xi}_t\right)$. We can then rewrite C_1 as

$$C_1=\frac{1}{T^2h^{3d/2}}\sum_{2\le t<s\le T}\Phi\left(\boldsymbol{\xi}_t,\boldsymbol{\xi}_s\right)$$

$$=\frac{1}{T^2h^{3d/2}}\sum_{2\le t<s\le T}\left[\Phi\left(\boldsymbol{\xi}_t,\boldsymbol{\xi}_s\right)-\Phi\left(\boldsymbol{\xi}_t\right)-\Phi\left(\boldsymbol{\xi}_s\right)+\Phi\right]$$

$$+\frac{T-1}{T^2h^{3d/2}}\sum_{t=2}^{T}\left[\Phi\left(\boldsymbol{\xi}_t\right)-\Phi\right]+\frac{T-1}{2Th^{3d/2}}\Phi$$

$$\equiv R_3+R_4+C_2, \tag{A.5}$$

where we further decompose

$$
\begin{aligned}
C_2 &= \frac{T-1}{Th^{3d/2}} \iiint \frac{a(\mathbf{x})K\left(\frac{\boldsymbol{\tau}-\mathbf{x}}{h}\right)^2}{g(\mathbf{x})} \left[1-|\varphi(\mathbf{u},t|\boldsymbol{\tau},\boldsymbol{\theta}_0)|^2\right] g(\boldsymbol{\tau})\, d\boldsymbol{\tau} d\mathbf{x} dW(\mathbf{u}) \\
&= \frac{T-1}{Th^{d/2}} \iiint \frac{a(h\mathbf{y}+\boldsymbol{\tau})K(\mathbf{y})^2}{g(h\mathbf{y}+\boldsymbol{\tau})} \left[1-|\varphi(\mathbf{u},t|\boldsymbol{\tau},\boldsymbol{\theta}_0)|^2\right] g(\boldsymbol{\tau})\, d\boldsymbol{\tau} d\mathbf{y} dW(\mathbf{u}) \\
&= \frac{1}{h^{d/2}} \iint \left[1-|\varphi(\mathbf{u},t|\boldsymbol{\tau},\boldsymbol{\theta}_0)|^2\right] a(\boldsymbol{\tau})\, d\boldsymbol{\tau} dW(\mathbf{u}) \int K(\mathbf{y})^2 d\mathbf{y} \\
&\quad + \left\{ \frac{T-1}{Th^{d/2}} \iiint \frac{a(h\mathbf{y}+\mathbf{x})K(\mathbf{y})^2}{g(h\mathbf{y}+\mathbf{x})} \left[1-|\varphi(\mathbf{u},t|\mathbf{x},\boldsymbol{\theta}_0)|^2\right] g(\mathbf{x})\, d\mathbf{x}\, d\mathbf{y}\, dW(\mathbf{u}) \right. \\
&\qquad \left. - \frac{1}{h^{d/2}} \iint \left[1-|\varphi(\mathbf{u},t|\mathbf{x},\boldsymbol{\theta}_0)|^2\right] a(\mathbf{x}) d\mathbf{x}\, dW(\mathbf{u}) \int K(\mathbf{y})^2 d\mathbf{y} \right\} \\
&\equiv \tilde{C} + R_{5,}
\end{aligned}
\tag{A.6}
$$

where $\varphi(\mathbf{u},t|\mathbf{x},\boldsymbol{\theta}) \equiv \varphi\left(\mathbf{u},t|\mathbf{X}_{t-1}=\mathbf{x},\boldsymbol{\theta}\right)$ and we have used the fact that $\mathrm{E}\left[|\mathbf{Z}_t(\mathbf{u},\boldsymbol{\theta}_0)|^2|\mathbf{X}_{t-1}\right] = 1-\left|\varphi\left(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}_0\right)\right|^2$.

We also decompose U_1 in a similar way. Define

$$
\begin{aligned}
\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r,\boldsymbol{\xi}_t) \equiv & \int \frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) K\left(\frac{\mathbf{X}_{r-1}-\mathbf{X}_{t-1}}{h}\right) \\
& \times \mathrm{Re}\left[\mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_r^*(\mathbf{u},\boldsymbol{\theta}_0)\right] dW(\mathbf{u}) \\
& + \int \frac{a\left(\mathbf{X}_{s-1}\right)}{g^2\left(\mathbf{X}_{s-1}\right)} K\left(\frac{\mathbf{X}_{r-1}-\mathbf{X}_{s-1}}{h}\right) K\left(\frac{\mathbf{X}_{t-1}-\mathbf{X}_{s-1}}{h}\right) \\
& \times \mathrm{Re}\left[\mathbf{Z}_r(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_t^*(\mathbf{u},\boldsymbol{\theta}_0)\right] dW(\mathbf{u}) \\
& + \int \frac{a\left(\mathbf{X}_{r-1}\right)}{g^2\left(\mathbf{X}_{r-1}\right)} K\left(\frac{\mathbf{X}_{t-1}-\mathbf{X}_{r-1}}{h}\right) K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{r-1}}{h}\right) \\
& \times \mathrm{Re}\left[\mathbf{Z}_t(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_s^*(\mathbf{u},\boldsymbol{\theta}_0)\right] dW(\mathbf{u}),
\end{aligned}
$$

$$
\begin{aligned}
\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r) \equiv \int \phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r,\boldsymbol{\xi}_t)\, dP(\boldsymbol{\xi}_t) = & \int \frac{a(\mathbf{x})}{g(\mathbf{x})} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) \\
& \times K\left(\frac{\mathbf{X}_{r-1}-\mathbf{x}}{h}\right) d\mathbf{x} \int \mathrm{Re}\left[\mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_r^*(\mathbf{u},\boldsymbol{\theta}_0)\right] dW(\mathbf{u}).
\end{aligned}
$$

We can rewrite U_1 as

$$U_1 = \frac{1}{3T^2h^{3d/2}} \sum_{t\neq s\neq r} [\phi(\boldsymbol{\xi}_s, \boldsymbol{\xi}_r, \boldsymbol{\xi}_t) - \phi(\boldsymbol{\xi}_s, \boldsymbol{\xi}_r) - \phi(\boldsymbol{\xi}_r, \boldsymbol{\xi}_t) - \phi(\boldsymbol{\xi}_s, \boldsymbol{\xi}_t)]$$

$$- \frac{2}{T^2h^{3d/2}} \sum_{s\neq r} \phi(\boldsymbol{\xi}_s, \boldsymbol{\xi}_r) + \frac{1}{Th^{3d/2}} \sum_{s\neq r} \phi(\boldsymbol{\xi}_s, \boldsymbol{\xi}_r)$$

$$\equiv R_6 - R_7 + 2\tilde{U}, \tag{A.7}$$

where we have used the fact that $\int \phi(\boldsymbol{\xi}_s, \boldsymbol{\xi}_r)\, dP(\boldsymbol{\xi}_r) = 0$. Proposition A.3 follows from the following seven lemmas.

LEMMA A.3. *Let R_1 be defined as in (A.4). Then $R_1 = o_P(1)$.*

LEMMA A.4. *Let R_2 be defined as in (A.4). Then $R_2 = o_P(1)$.*

LEMMA A.5. *Let R_3 be defined as in (A.5). Then $R_3 = o_P(1)$.*

LEMMA A.6. *Let R_4 be defined as in (A.5). Then $R_4 = o_P(1)$.*

LEMMA A.7. *Let R_5 be defined as in (A.6). Then $R_5 = o_P(1)$.*

LEMMA A.8. *Let R_6 be defined as in (A.7). Then $R_6 = o_P(1)$.*

LEMMA A.9. *Let R_7 be defined as in (A.7). Then $R_7 = o_P(1)$.*

Proof of Lemma A.3. It follows immediately from Assumptions A.4 and A.5. ■

Proof of Lemma A.4. We write

$$\Psi(\boldsymbol{\xi}_t, \boldsymbol{\xi}_s) = \frac{1}{T^2h^{3d/2}} \int \frac{a(\mathbf{X}_{t-1})}{g^2(\mathbf{X}_{t-1})} K(\mathbf{0}) K\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right)$$

$$\times \operatorname{Re}\left[\mathbf{Z}_s(\mathbf{u}, \boldsymbol{\theta}_0) \mathbf{Z}_t^*(\mathbf{u}, \boldsymbol{\theta}_0)\right] dW(\mathbf{u})$$

$$+ \frac{1}{T^2h^{3d/2}} \int \frac{a(\mathbf{X}_{s-1})}{g^2(\mathbf{X}_{s-1})} K(\mathbf{0}) K\left(\frac{\mathbf{X}_{t-1} - \mathbf{X}_{s-1}}{h}\right)$$

$$\times \operatorname{Re}\left[\mathbf{Z}_t(\mathbf{u}, \boldsymbol{\theta}_0) \mathbf{Z}_s^*(\mathbf{u}, \boldsymbol{\theta}_0)\right] dW(\mathbf{u}).$$

We can rewrite $R_2 = \frac{1}{2T^2h^{3d/2}} \sum_{2\leq t<s\leq T} \Psi(\boldsymbol{\xi}_t, \boldsymbol{\xi}_s)$, where $\int \Psi(\boldsymbol{\xi}_t, \boldsymbol{\xi}_s)\, dP(\boldsymbol{\xi}_s) = 0$. By Lemma A(ii) of Hjellvik et al. (1998), we have $\mathrm{E}R_2^2 \leq \frac{C}{T^2h^{3d}} \left[\mathrm{E}\left|\Psi(\boldsymbol{\xi}_t, \boldsymbol{\xi}_s)\right|^{2(1+\delta)}\right]^{1/(1+\delta)} \times \sum_{j=1}^{T} j\beta^{\delta/(1+\delta)}(j) \leq CT^{-2}h^{-(3d-d/(1+\delta))}$, where we have used Assumption A.1 and change of variable. Then Lemma A.4 follows from Chebyshev's inequality and Assumption A.5. ■

Proof of Lemma A.5. By Lemma A(ii) of Hjellvik et al. (1998), we have $\mathrm{E}R_3^2 \leq \frac{C}{T^2h^{3d}} [\mathrm{E}|\Phi(\boldsymbol{\xi}_1, \boldsymbol{\xi}_s) - \Phi(\boldsymbol{\xi}_1) - \Phi(\boldsymbol{\xi}_s) + \Phi|^{2(1+\delta)}]^{1/(1+\delta)} \sum_{j=1}^{T} j\beta^{\delta/(1+\delta)}(j) \leq$

$CT^{-2}h^{-(3d-d/(1+\delta))}$, where we have used Assumption A.1 and change of variable. Then Lemma A.5 follows from Chebyshev's inequality and Assumption A.5. ■

Proof of Lemma A.6. We have $\mathrm{E}R_4^2 \le \frac{1}{T^2h^{3d}}\sum_{t=2}^{T}\mathrm{var}[\Phi(\boldsymbol{\xi}_t)] + \frac{2}{Th^{3d}}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\mathrm{cov}\left[\Phi(\boldsymbol{\xi}_1),\Phi\left(\boldsymbol{\xi}_{1+j}\right)\right] \le \frac{C}{Th^d} + \frac{2}{Th^d}\sum_{j=2}^{T-1}\beta^{\delta/(1+\delta)}(j) = O\left(T^{-1}h^{-d}\right)$, where we have used the mixing inequality (e.g., Fan and Yao, 2003, p. 72) and Assumption A.1. Then Lemma A.6 follows from Chebyshev's inequality and Assumption A.5. ■

Proof of Lemma A.7. By a second-order Taylor expansion and the continuity of $g(\cdot)$ and $a(\cdot)$ over $\mathbf{G}$. ■

Proof of Lemma A.8. First note that for any $1\le s<r<t\le T$ and $\delta>0$, we have $\int|\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r,\boldsymbol{\xi}_t)|^{2(1+\delta)}dP \le Ch^{2d}$, where P denotes any one of the probability measures in the set $\{P(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r,\boldsymbol{\xi}_t), P(\boldsymbol{\xi}_s)P(\boldsymbol{\xi}_r,\boldsymbol{\xi}_t), P(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)P(\boldsymbol{\xi}_t), P(\boldsymbol{\xi}_s)P(\boldsymbol{\xi}_r)P(\boldsymbol{\xi}_t)\}$. By Lemma A(i) of Hjellvik et al. (1998), we have $\mathrm{E}R_5^2 = O(T^{-1}h^{-(d+3d\delta)/(1+\delta)})$. Then Lemma A.8 follows from Chebyshev's inequality and Assumption A.5. ■

Proof of Lemma A.9. It follows by the fact that $\tilde{U} = O_P(1)$, as is implied by Proposition A.4 (the proof of Proposition A.4 does not depend on Lemma A.9). ■

Proof of Proposition A.4. We apply the Hjellvik et al. (1998) central limit theorem for degenerate U-statistics, which states that $\sigma_T^{-1}\sum_{s<r}\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r) \xrightarrow{d} N(0,1)$ if for some $\delta>0$, $\sum_{j=1}^{\infty} j^2\beta^{\delta/(1+\delta)}(j)<\infty$ and

$$\max\frac{1}{\sigma_T^2}\left\{T^2\left[M_{T1}^{1/(1+\delta)}+M_{T5}^{1/2(1+\delta)}+M_{T6}^{1/2}\right],\right.$$
$$\left.T^{3/2}\left[M_{T2}^{1/2(1+\delta)}+M_{T3}^{1/2}+M_{T4}^{1/2(1+\delta)}\right]\right\}\to 0, \tag{A.8}$$

as $T\to\infty$, where $\sigma_T^2 = \sum_{1\le s<r\le T}\mathrm{Var}[\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)]$ and

$$M_{T1} = \max_{1<s<r\le T}\max\left\{\mathrm{E}|\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)|^{1+\delta},\right.$$
$$\left.\int|\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)|^{1+\delta}dP(\boldsymbol{\xi}_1)dP(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)\right\},$$

$$M_{T2} = \max_{1<s<r\le T}\max\left\{\mathrm{E}|\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)|^{2(1+\delta)},\right.$$
$$\int|\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)|^{2(1+\delta)}dP(\boldsymbol{\xi}_1)dP(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r),$$
$$\int|\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)|^{2(1+\delta)}dP(\boldsymbol{\xi}_1,\boldsymbol{\xi}_s)dP(\boldsymbol{\xi}_r),$$
$$\left.\int|\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)|^{2(1+\delta)}dP(\boldsymbol{\xi}_1)dP(\boldsymbol{\xi}_s)dP(\boldsymbol{\xi}_r)\right\},$$

$$M_{T3} = \max_{1\le s<r\le T}\mathrm{E}|\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)|^2,$$

$$M_{T4} = \max_{\substack{1<s,r,t\leq T\\ s,r,t\,\text{different}}} \left\{ \max_P \int |\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_t)|^{2(1+\delta)}\,dP \right\},$$

$$M_{T5} = \max\max_{1<s<r} \left\{ \mathrm{E}\left| \int \phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_s)\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\,dP(\boldsymbol{\xi}_1) \right|^{2(1+\delta)}, \right.$$

$$\left. \int \left| \int \phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_s)\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\,dP(\boldsymbol{\xi}_1) \right|^{2(1+\delta)} dP(\boldsymbol{\xi}_s)\,dP(\boldsymbol{\xi}_r) \right\},$$

$$M_{T6} = \max_{1<s<r} \mathrm{E}\left| \int \phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_s)\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\,dP(\boldsymbol{\xi}_1) \right|^2,$$

and the maximization over P in the equation for M_{T4} is taken over the four probability measures $P(\boldsymbol{\xi}_1,\boldsymbol{\xi}_s,\boldsymbol{\xi}_r,\boldsymbol{\xi}_t)$, $P(\boldsymbol{\xi}_1)P(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r,\boldsymbol{\xi}_t)$, $P(\boldsymbol{\xi}_1)P(\boldsymbol{\xi}_{s_1})P(\boldsymbol{\xi}_{s_2},\boldsymbol{\xi}_{s_3})$, and $P(\boldsymbol{\xi}_1)P(\boldsymbol{\xi}_s)P(\boldsymbol{\xi}_r)P(\boldsymbol{\xi}_t)$, where (s_1,s_2,s_3) is the permutation of (s,r,t) in ascending order.

First we calculate the asymptotic variance. First, we have

$$\begin{aligned}
\sigma_0^2 &= \int \phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)^2\,dP(\boldsymbol{\xi}_s)\,dP(\boldsymbol{\xi}_r)\\
&= \mathrm{E}\left\{ \left[\int \mathrm{Re}\mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_r^*(\mathbf{u},\boldsymbol{\theta}_0)\,dW(\mathbf{u}) \right]^2 \right.\\
&\quad \times \left. \left[\int \frac{K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right)K\left(\frac{\mathbf{X}_{r-1}-\mathbf{x}}{h}\right)a(\mathbf{x})}{g(\mathbf{x})}\,d\mathbf{x} \right]^2 \right\}\\
&= h^{3d} \iiint |\varphi(\mathbf{u}_1+\mathbf{u}_2,t|\mathbf{x},\boldsymbol{\theta}_0) - \varphi(\mathbf{u}_1,t|\mathbf{x},\boldsymbol{\theta}_0)\varphi(\mathbf{u}_2,t|\mathbf{x},\boldsymbol{\theta}_0)|^2\\
&\quad \times a^2(\mathbf{x})d\mathbf{x}\,dW(\mathbf{u}_1)\,dW(\mathbf{u}_2) \left\{ \int \left[\int K(\boldsymbol{\tau})K(\boldsymbol{\tau}+\boldsymbol{\eta})\,d\tau \right]^2 d\boldsymbol{\eta} \right\} + o\left(h^{3d}\right),
\end{aligned}$$

where $\varphi(\mathbf{u},t|\mathbf{x},\boldsymbol{\theta}) \equiv \varphi\left(\mathbf{u},t|\mathbf{X}_{t-1}=\mathbf{x},\boldsymbol{\theta}\right)$ and we have used Assumptions A.4 and A.5 and the fact that

$$\mathrm{E}[\mathbf{Z}_1(\mathbf{u}_1,\boldsymbol{\theta}_0)\mathbf{Z}_1(\mathbf{u}_2,\boldsymbol{\theta}_0)|\mathbf{X}_0] = \varphi(\mathbf{u}_1+\mathbf{u}_2,1|\mathbf{X}_0,\boldsymbol{\theta}_0) - \varphi(\mathbf{u}_1,1|\mathbf{X}_0,\boldsymbol{\theta}_0)\varphi(\mathbf{u}_2,1|\mathbf{X}_0,\boldsymbol{\theta}_0).$$

It follows from the proof of Theorem A in Hjellvik et al. (1998) that $\sigma_T^2 = \frac{T^2}{2}\sigma_0^2[1+o(1)] = O(T^2h^{3d})$.

Now we verify condition (A.8). Observe that

$$\begin{aligned}
&\mathrm{E}_{s<r}|\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\phi(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)|^{1+\delta}\\
&\quad = \mathrm{E}\left| \iiiint \frac{\mathrm{Re}\left[\mathbf{Z}_1(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_r^*(\mathbf{v},\boldsymbol{\theta}_0)\right]}{g(\mathbf{x})} K\left(\frac{\mathbf{X}_0-\mathbf{x}}{h}\right)K\left(\frac{\mathbf{X}_{r-1}-\mathbf{x}}{h}\right)a(\mathbf{x})\right.
\end{aligned}$$

$$\times \frac{\operatorname{Re}\left[\mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_r^*(\mathbf{v},\boldsymbol{\theta}_0)\right]}{g(\mathbf{y})} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{y}}{h}\right)$$

$$\left.\times K\left(\frac{\mathbf{X}_{r-1}-\mathbf{y}}{h}\right) a(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y}\,dW(\mathbf{u})\,dW(\mathbf{v})\right|^{1+\delta}$$

$$= h^{2d(1+\delta)}\mathrm{E}\left|\iiiint \frac{\operatorname{Re}\left[\mathbf{Z}_1(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_r^*(\mathbf{v},\boldsymbol{\theta}_0)\right]}{g(\mathbf{X}_0+\tau h)} K(\boldsymbol{\tau}) K\left(\boldsymbol{\tau}+\frac{\mathbf{X}_0-\mathbf{X}_{r-1}}{h}\right)\right.$$

$$\times a(\mathbf{X}_0+\boldsymbol{\tau}h)\frac{\operatorname{Re}\left[\mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_r^*(\mathbf{v},\boldsymbol{\theta}_0)\right]}{g\left(\mathbf{X}_{s-1}+\boldsymbol{\omega}h\right)} K(\boldsymbol{\omega}) K\left(\boldsymbol{\omega}+\frac{\mathbf{X}_{s-1}-\mathbf{X}_{r-1}}{h}\right)$$

$$\left.\times\, a\left(\mathbf{X}_{s-1}+\boldsymbol{\omega}h\right) d\boldsymbol{\tau}\,d\boldsymbol{\omega}\,dW(\mathbf{u})\,dW(\mathbf{v})\right|^{1+\delta} = O\left(h^{2d(2+\delta)}\right),$$

where we have used Assumptions A.4 and A.5. By similar arguments, we have $M_{T1} = O\left(h^{2d(2+\delta)}\right)$, $M_{T2} = O\left(h^{2d(3+2\delta)}\right)$, $M_{T3} = O\left(h^{6d}\right)$, $M_{T4} = O\left(h^{2d(3+2\delta)}\right)$. On the other hand, we have

$$\mathrm{E}\left|\int \phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_s)\phi(\boldsymbol{\xi}_1,\boldsymbol{\xi}_r)\,dP(\boldsymbol{\xi}_1)\right|^{2(1+\delta)}$$

$$= \mathrm{E}\left|\int\cdots\int \frac{\operatorname{Re}\left[\mathbf{z}_1(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_s^*(\mathbf{v},\boldsymbol{\theta}_0)\right]}{g(\mathbf{x})} K\left(\frac{\mathbf{x}_0-\mathbf{x}}{h}\right) K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) a(\mathbf{x})\right.$$

$$\times \frac{\operatorname{Re}\left[z_1(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_r^*(\mathbf{v},\boldsymbol{\theta}_0)\right]}{g(\mathbf{y})} K\left(\frac{\mathbf{x}_0-\mathbf{y}}{h}\right) K\left(\frac{\mathbf{X}_{r-1}-\mathbf{y}}{h}\right)$$

$$\left.\times\, a(\mathbf{y})\,g(\mathbf{x}_0,\mathbf{x}_1)\,d\mathbf{x}_0\,d\mathbf{x}_1\,d\mathbf{x}\,d\mathbf{y}\,dW(\mathbf{u})\,dW(\mathbf{v})\right|^{2(1+\delta)}$$

$$= h^{4d(1+\delta)}\mathrm{E}\left|\int\cdots\int \frac{\operatorname{Re}\left[\mathbf{z}_1(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_s^*(\mathbf{v},\boldsymbol{\theta}_0)\right]}{g(\mathbf{x}_0+\boldsymbol{\tau}h)} K(\boldsymbol{\tau}) K\left(\boldsymbol{\tau}+\frac{\mathbf{x}_0-\mathbf{X}_{s-1}}{h}\right)\right.$$

$$\times\, a(\mathbf{x}_0+\boldsymbol{\tau}h)\frac{\operatorname{Re}\left[z_1(\mathbf{u},\boldsymbol{\theta}_0)\mathbf{Z}_r^*(\mathbf{v},\boldsymbol{\theta}_0)\right]}{g(\mathbf{x}_0+\boldsymbol{\omega}h)} K(\boldsymbol{\omega}) K\left(\boldsymbol{\omega}+\frac{\mathbf{x}_0-\mathbf{X}_{r-1}}{h}\right)$$

$$\left.\times\, a(\mathbf{x}_0+\boldsymbol{\omega}h)\,g(\mathbf{x}_0,\mathbf{x}_1)\,d\mathbf{x}_0\,d\mathbf{x}_1\,d\boldsymbol{\tau}\,d\boldsymbol{\omega}\,dW(\mathbf{u})\,dW(\mathbf{v})\right|^{1+\delta}$$

$$= O\left(h^{7d+6d\delta}\right),$$

where we have used Assumptions A.4 and A.5. By similar arguments, we have $M_{T5} = O\left(h^{7d+6d\delta}\right)$ and $M_{T6} = O\left(h^{7d}\right)$. Hence condition (A.8) holds, and $\tilde{U} \xrightarrow{d} N(0,\tilde{D}/2)$ by the Hjellvik et al. (1998) central limit theorem. ■

Proof of Theorem 2. Given Theorem 1, it suffices to show

$$\frac{1}{\sqrt{2\hat{D}}}\sum_{t=2}^{T}\int\left[\hat{h}^{d/2}\left|\hat{m}_{\hat{h}}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right|^2-h^{d/2}\left|\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right|^2\right]$$
$$\times a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})=o_P(1), \quad \textbf{(A.9)}$$

and $\hat{C}_{\hat{h}}-\hat{C}=o_P(1)$. For space, we focus on the proof of (A.9); the proof for $\hat{C}_{\hat{h}}-\hat{C}=o_P(1)$ is straightforward.

To show (A.9), we first decompose

$$\sum_{t=2}^{T}\int\left[\hat{h}^{d/2}\left|\hat{m}_{\hat{h}}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right|^2-h^{d/2}\left|\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right|^2\right]a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})$$
$$=\left(\hat{h}^{d/2}-h^{d/2}\right)\sum_{t=2}^{T}\int\left|\hat{m}_{\hat{h}}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})$$
$$+h^{d/2}\sum_{t=2}^{T}\int\left|\hat{m}_{\hat{h}}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})-\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})$$
$$+2\mathrm{Re}\left[h^{d/2}\sum_{t=2}^{T}\int\left[\hat{m}_{\hat{h}}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})-\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right]\right.$$
$$\left.\times\,\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})^{*}a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})\right]\equiv\hat{B}_1+\hat{B}_2+\hat{B}_3.$$

It is easy to see that under the conditions of Theorem 2, $\hat{B}_1=(\hat{h}^{d/2}-h^{d/2})/h^{d/2}\sum_{t=2}^{T}h^{d/2}\int\left|\hat{m}_{\hat{h}}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})=o_P(1)$ if $\hat{B}_2=o_P(1)$ and $\hat{B}_3=o_P(1)$. Therefore, we focus on Theorems A.3 and A.4.

THEOREM A.3. *Under the conditions of Theorem 2,* $\hat{B}_2\xrightarrow{p}0$.

THEOREM A.4. *Under the conditions of Theorem 2,* $\hat{B}_3\xrightarrow{p}0$.

We first state a lemma.

LEMMA A.10. *Under the conditions of Theorem 2,* $\hat{W}_{\hat{h}}(t)=\frac{1}{Th^d g(\mathbf{x})}K(t)\left[1+o_P\left(h^{d/2}\right)\right]$ *uniformly in* $\mathbf{x}\in\mathbf{G}$.

Proof of Lemma A.10. First we note that $\hat{h}^{-d}=h^{-d}+h^{-d}\left((h-\hat{h})/\hat{h}\right)^d=h^{-d}\left[1+o_P\left(h^{d/2}\right)\right]$, where we have used the facts that $\left(\hat{h}-h\right)/h=o_P\left(h^{d/2}\right)$ and $h=cT^{-\lambda}$. For the element of $\mathbf{S}_T$, we have

$$\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'K_{\hat{h}}\left(\mathbf{X}_t-\mathbf{x}\right)-g\left(\mathbf{x}\right)\mu_j\right|$$
$$\leq\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'K_h\left(\mathbf{X}_t-\mathbf{x}\right)-g\left(\mathbf{x}\right)\mu_j\right|$$

$$+\left|\frac{\hat{h}^{-d}-h^{-d}}{T}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'K\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\right|$$

$$+\left|T^{-1}\hat{h}^{-d}\sum_{t=1}^{T}\left[\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'K\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)\right.\right.$$

$$\left.\left.-\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'K\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\right]\right|\equiv S_1+S_2+S_3,$$

where $\mu_2 \equiv \int_{\mathbb{R}^d}\mathbf{u}\mathbf{u}'K(\mathbf{u})d\mathbf{u}$. The convergence of other elements of $\mathbf{S}_T$ can be shown similarly. For the first term, we have $S_1 = o_P(1)$ uniformly in $\mathbf{x}\in\mathbf{G}$ by Proposition 6.2 of Fan and Yao (2003). For the second term, we have

$$\sup_{\mathbf{x}\in\mathbf{G}} S_2 \le \left|\frac{\hat{h}^{-d}-h^{-d}}{h^{-d}}\right|\sup_{\mathbf{x}\in\mathbf{G}}\left|T^{-1}h^{-d}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'K\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\right.$$

$$\left.-\mathrm{E}\left[T^{-1}h^{-d}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'K\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\right]\right|+\left|\frac{\hat{h}^{-d}-h^{-d}}{h^{-d}}\right|$$

$$\times\sup_{\mathbf{x}\in\mathbf{G}}\left|\mathrm{E}\left[T^{-1}h^{-d}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'K\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\right]\right|$$

$$= o_P\left(h^{d/2}\right)o_P(1)+o_P\left(h^{d/2}\right)O_P(1)=o_P\left(h^{d/2}\right),$$

where we have used Proposition 1 of Masry (1996). For the last term, we have

$$\sup_{\mathbf{x}\in\mathbf{G}} S_2 = \sup_{\mathbf{x}\in\mathbf{G}}\left|T^{-1}\hat{h}^{-d}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'K\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)\left[1-\left(\frac{\hat{h}}{h}\right)^j\right]\right|$$

$$+\sum_{s=1}^{2}\frac{1}{s}\sup_{\mathbf{x}\in\mathbf{G}}\left|T^{-1}\hat{h}^{-d}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'\right.$$

$$\left.\times K^{(s)}\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)\left(\frac{\hat{h}-h}{h}\right)^s\right|$$

$$+\frac{1}{6}\sup_{\mathbf{x}\in\mathbf{G}}\left|T^{-1}\hat{h}^{-d}\sum_{t=1}^{T}\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_t-\mathbf{x}}{\hat{h}}\right)'\right.$$

$$\left.\times K^{(3)}\left(\frac{\mathbf{X}_t-\mathbf{x}}{\bar{h}}\right)\left(\frac{\hat{h}-h}{\bar{h}}\right)^3\right|$$

$$= O_P(1)o_P\left(h^{d/2}\right)+O_P(1)o_P\left(h^{d/2}\right)+O_P\left(h^{-d}\right)o_P\left(h^{3d/2}\right)=o_P\left(h^{d/2}\right),$$

where $K^{(s)}\left((\mathbf{X}_t-\mathbf{x})/h\right)\equiv h^s\left(\partial^s/\partial h^s\right)K\left((\mathbf{X}_t-\mathbf{x})/h\right)$ can be viewed as a second-order kernel function and $\bar{h}$ is between h and $\hat{h}$. Hence Lemma A.10 follows from the same proof as Proposition 6.2 of Fan and Yao (2003). ■

Proof of Theorem A.3. By Lemma A.10, the Taylor expansion, change of variable, and Assumptions A.2–A.6, we have

$$
\begin{aligned}
\hat{B}_2 &= h^{d/2}\sum_{t=2}^{T}\int\Bigg|\frac{1}{T\hat{h}^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{\hat{h}}\right)\mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\\
&\qquad -\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\Bigg|^2\\
&\quad\times a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})\left[1+o_P\left(h^{d/2}\right)\right]\\
&= h^{d/2}\sum_{t=2}^{T}\int\Bigg|\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{\hat{h}}\right)\mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\left(\frac{h^d-\hat{h}^d}{\hat{h}^d}\right)\Bigg|^2\\
&\quad\times a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})\left[1+o_P\left(h^{d/2}\right)\right]\\
&\quad+h^{d/2}\sum_{t=2}^{T}\int\Bigg|\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}\sum_{v=1}^{2}K^{(v)}\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\\
&\qquad\times\mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\left(\frac{\hat{h}-h}{h}\right)^s\Bigg|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})\left[1+o_P\left(h^{d/2}\right)\right]\\
&\quad+h^{d/2}\sum_{t=2}^{T}\int\Bigg|\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K^{(3)}\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{\bar{h}}\right)\mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\left(\frac{\hat{h}-h}{\bar{h}}\right)^3\Bigg|^2\\
&\quad\times a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})\left[1+o_P\left(h^{d/2}\right)\right]\\
&= O_P\left(h^{-d/2}\right)o_P\left(h^{d^2}\right)+O_P\left(h^{-d/2}\right)o_P\left(h^{2d}\right)+o_P\left(h^{3d/2}\right)=o_P(1).
\end{aligned}
$$
■

Proof of Theorem A.4. By Lemma A.10, we have

$$
\begin{aligned}
&h^{d/2}\sum_{t=2}^{T}\int\left|\left[\hat{m}_{\hat{h}}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})-\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right]\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})^*\right|a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})\\
&\quad\le h^{d/2}\sum_{t=2}^{T}\int\left|\frac{1}{T\hat{h}^d g\left(\mathbf{X}_{t-1}\right)}\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{\hat{h}}\right)\mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\right.
\end{aligned}
$$

$$- \frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)} \sum_{s=2}^{T} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) \mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\Bigg|$$

$$\times \left|\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)} \sum_{s=2}^{T} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) \mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\right|$$

$$\times a\left(\mathbf{X}_{t-1}\right) dW(\mathbf{u}) \left[1+o_P\left(h^{d/2}\right)\right]$$

$$\leq \hat{B}_2^{1/2} \left[h^{d/2} \sum_{t=2}^{T} \int \left|\frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)} \sum_{s=2}^{T} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) \mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\right|^2\right.$$

$$\left. \times a\left(\mathbf{X}_{t-1}\right) dW(\mathbf{u})\right]^{1/2} \left[1+o_P\left(h^{d/2}\right)\right] = o_P(1),$$

where we have used Theorem A.3 and $h^{d/2}\sum_{t=2}^{T}\int\left|\frac{1}{Th^d g(\mathbf{X}_{t-1})}\sum_{s=2}^{T} K((\mathbf{X}_{s-1} - \mathbf{X}_{t-1})/h)\mathbf{Z}_s\left(\mathbf{u},\hat{\boldsymbol{\theta}}\right)\right|^2 a\left(\mathbf{X}_{t-1}\right) dW(\mathbf{u}) = O_P\left(h^{-d/2}\right)$ from Theorem A.1 and Proposition A.3. ■

Proof of Theorem 3(i). The proof of Theorem 3(i) consists of the proofs of Theorems A.5–A.7.

THEOREM A.5. *Under the conditions of Theorem 3,* $T^{-1}h^{-d/2}\left(\hat{M}-\tilde{M}\right) \xrightarrow{p} 0.$

THEOREM A.6. *Put* $M \equiv \frac{1}{\sqrt{2\tilde{D}}}\left[\iint \frac{1}{Th^{3d/2}}\left|\sum_{s=2}^{T} K\left((\mathbf{X}_{s-1}-\mathbf{x})/h\right)\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}^*\right)\right|^2 \frac{a(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x}dW(\mathbf{u}) - \tilde{C}\right]$. *Then under the conditions of Theorem 2,* $T^{-1}h^{-d/2}\left(M-\tilde{M}\right) \xrightarrow{p} 0.$

THEOREM A.7. *Under the conditions of Theorem 3,* $T^{-1}h^{-d/2}M \xrightarrow{p} (2D)^{-1/2} \iint \left|m\left(\mathbf{u},\mathbf{x},\boldsymbol{\theta}^*\right)\right|^2 a(\mathbf{x})\, g(\mathbf{x}) d\mathbf{x} dW(\mathbf{u})$.

Proof of Theorem A.5. It suffices to show that $\hat{C}-\tilde{C} = o_P(1)$, $\hat{D}-\tilde{D} = o_P(1)$, and

$$\frac{1}{T}\sum_{t=2}^{T}\int\left[\left|\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\hat{\boldsymbol{\theta}})\right|^2 - \left|\hat{m}(\mathbf{u},\mathbf{X}_{t-1},\boldsymbol{\theta}_0)\right|^2\right] a\left(\mathbf{X}_{t-1}\right) dW(\mathbf{u}) = o_P(1). \tag{A.10}$$

Because the proofs for $\hat{C}-\tilde{C} = o_P(1)$ and $\hat{D}-\tilde{D} = o_P(1)$ are tedious but straightforward, we focus on the proof of (A.10). From (A.10), the Cauchy–Schwarz inequality, and $\sum_{t=2}^{T}\int \frac{a(\mathbf{X}_{t-1})}{T^3h^{2d}g^2(\mathbf{X}_{t-1})}\left|\sum_{s=2}^{T} K((\mathbf{X}_{s-1}-\mathbf{X}_{t-1})/h)\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}_0\right)\right|^2 \times dW(\mathbf{u}) = O_P(h^{d/2})$ implied by Theorem A.4 (the proof of Theorem A.4 does not depend on Theorem A.3), it suffices to show that $T^{-1}h^{-d/2}\hat{A}_1 \xrightarrow{p} 0$, where $\hat{A}_1$ is defined as in (A.2). The proof of $T^{-1}h^{-d/2}\hat{A}_1 \xrightarrow{p} 0$ follows from Lemmas A.10 and A.11. ■

LEMMA A.11. *Under the conditions of Theorem 3,* $\frac{1}{T}\sum_{t=2}^{T}\int\left|\hat{A}_{11}(\mathbf{u},\mathbf{X}_{t-1})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u}) = o_P(1)$.

LEMMA A.12. *Under the conditions of Theorem 3,* $\frac{1}{T}\sum_{t=2}^{T}\int\left|\hat{A}_{12}(\mathbf{u},\mathbf{X}_{t-1})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u}) = o_P(1)$.

Proof of Lemma A.11. By Assumptions A.2, A.4, and A.5, Markov's inequality, and change of variable, we decompose

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=2}^{T}\int\left|\hat{A}_{11}(\mathbf{u},\mathbf{X}_{t-1})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})\\
&\quad=\sum_{t=2}^{T}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^3h^{2d}g^2\left(\mathbf{X}_{t-1}\right)}K\left(\mathbf{0}\right)^2\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}^*\right)\\
&\qquad\times\frac{\partial}{\partial\boldsymbol{\theta}}\varphi^*\left(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}^*\right)dW(\mathbf{u})\left\|\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right)\right\|^2\\
&\qquad+C\sum_{s>t}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^3h^{2d}g^2\left(\mathbf{X}_{t-1}\right)}K\left(\mathbf{0}\right)K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\\
&\qquad\times\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left(\mathbf{u},t|\mathbf{X}_{t-1},\boldsymbol{\theta}^*\right)\frac{\partial}{\partial\boldsymbol{\theta}}\varphi^*\left(\mathbf{u},s|\mathbf{X}_{s-1},\boldsymbol{\theta}^*\right)dW(\mathbf{u})\left\|\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right)\right\|^2\\
&\qquad+C\sum_{s>t}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^3h^{2d}g^2\left(\mathbf{X}_{t-1}\right)}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)^2\\
&\qquad\times\left|\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left(\mathbf{u},s|\mathbf{X}_{s-1},\boldsymbol{\theta}^*\right)\right|^2 dW(\mathbf{u})\left\|\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right)\right\|^2\\
&\qquad+C\sum_{s'>s>t}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{T^3h^{2d}g^2\left(\mathbf{X}_{t-1}\right)}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)K\left(\frac{\mathbf{X}_{s'-1}-\mathbf{X}_{t-1}}{h}\right)\\
&\qquad\times\frac{\partial}{\partial\boldsymbol{\theta}}\varphi\left(\mathbf{u},s|\mathbf{X}_{s-1},\boldsymbol{\theta}^*\right)\\
&\qquad\times\frac{\partial}{\partial\boldsymbol{\theta}}\varphi^*\left(\mathbf{u},s|\mathbf{X}_{s'-1},\boldsymbol{\theta}^*\right)dW(\mathbf{u})\left\|\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right)\right\|^2\\
&\quad=o_P(T^{-2}h^{-2d}+T^{-1}h^{-d}+T^{-1}h^{-d})+o_P(1)=o_P(1).
\end{aligned}
$$

■

Proof of Lemma A.12. By Assumptions A.2–A.5, change of variable, and the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=2}^{T}\int\left|\hat{A}_{12}(\mathbf{u},\mathbf{X}_{t-1})\right|^2 a\left(\mathbf{X}_{t-1}\right)dW(\mathbf{u})\\
&\quad\leq\sum_{t=2}^{T}\int\frac{a\left(\mathbf{X}_{t-1}\right)}{2T^3h^{2d}g^2\left(\mathbf{X}_{t-1}\right)}\left|\sum_{s=2}^{T}K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\right|^2
\end{aligned}
$$

$$\times \left[\sup_{\theta \in \boldsymbol{\Theta}} \left\| \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \varphi\left(\mathbf{u}, t | \mathbf{X}_{t-1}, \boldsymbol{\theta}^*\right) \right\| \right]^2 dW(\mathbf{u}) \left\| \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \right\|^4$$

$$\leq C \sum_{t=2}^{T} \frac{a\left(\mathbf{X}_{t-1}\right) K^2(\mathbf{0})}{T^3 h^{2d} g^2\left(\mathbf{X}_{t-1}\right)} + C \sum_{s>t} \frac{a\left(\mathbf{X}_{t-1}\right) K^2\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right)}{T^3 h^{2d} g^2\left(\mathbf{X}_{t-1}\right)}$$

$$+ C \sum_{s>t} \frac{a\left(\mathbf{X}_{t-1}\right) K(0) K\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right)}{T^3 h^{2d} g^2\left(\mathbf{X}_{t-1}\right)}$$

$$+ C \sum_{s'>s>t} \int \frac{a\left(\mathbf{X}_{t-1}\right)}{T^3 h^{2d} g^2\left(\mathbf{X}_{t-1}\right)} K\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right) K\left(\frac{\mathbf{X}_{s'-1} - \mathbf{X}_{t-1}}{h}\right)$$

$$= o_P(T^{-2} h^{-2d} + T^{-1} h^{-d} + T^{-1} h^{-d}) + o_P(1) = o_P(1). \qquad \blacksquare$$

Proof of Theorem A.6. To show $T^{-1} h^{-\frac{d}{2}} \left(M - \tilde{M}\right) \xrightarrow{p} 0$, it suffices to show

$$R \equiv \frac{1}{T} \sum_{t=2}^{T} \int \left[\left| \frac{1}{T h^d g\left(\mathbf{X}_{t-1}\right)} \sum_{s=2}^{T} K\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right) \mathbf{Z}_s\left(\mathbf{u}, \boldsymbol{\theta}^*\right) \right|^2 a\left(\mathbf{X}_{t-1}\right) \right.$$

$$\left. - \int \left| \frac{1}{T h^d} \sum_{s=2}^{T} K\left(\frac{\mathbf{X}_{s-1} - \mathbf{x}}{h}\right) \mathbf{Z}_s\left(\mathbf{u}, \boldsymbol{\theta}^*\right) \right|^2 \frac{a(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x} \right] dW(\mathbf{u}) = o_P(1),$$

given Proposition 6.2 in Fan and Yao (2003, p. 236) and $\tilde{D} = O_P(1)$.

We first decompose

$$R_A = \frac{1}{T^3 h^{2d}} \sum_{t=2}^{T} \int \left| \mathbf{Z}_t\left(\mathbf{u}, \boldsymbol{\theta}^*\right) \right|^2 \left[\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)} K(\mathbf{0})^2 \right.$$

$$\left. - \int \frac{a(\mathbf{x})}{g(\mathbf{x})} K\left(\frac{\mathbf{X}_{t-1} - \mathbf{x}}{h}\right)^2 d\mathbf{x} \right] dW(\mathbf{u})$$

$$+ \frac{1}{T^3 h^{2d}} \sum_{t=2}^{T} \sum_{s \neq t} \int \left| \mathbf{Z}_s\left(\mathbf{u}, \boldsymbol{\theta}^*\right) \right|^2 \left[\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)} K\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right)^2 \right.$$

$$\left. - \int \frac{a(\mathbf{x})}{g(\mathbf{x})} K\left(\frac{\mathbf{X}_{s-1} - \mathbf{x}}{h}\right)^2 d\mathbf{x} \right] dW(\mathbf{u})$$

$$+ \frac{1}{T^3 h^{2d}} \sum_{t=2}^{T} \sum_{s \neq t} \int \operatorname{Re}\left[\mathbf{Z}_s\left(\mathbf{u}, \boldsymbol{\theta}^*\right) \mathbf{Z}_t^*\left(\mathbf{u}, \boldsymbol{\theta}^*\right)\right]$$

$$\times \left[\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)} K(\mathbf{0}) K\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right) \right.$$

$$
\begin{aligned}
&\quad - \int \frac{a(\mathbf{x})}{g(\mathbf{x})} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) K\left(\frac{\mathbf{X}_{t-1}-\mathbf{x}}{h}\right) d\mathbf{x}\Bigg] dW(\mathbf{u}) \\
&+ \frac{1}{T^3 h^{2d}} \sum_{t\neq s\neq r}^{T} \int \operatorname{Re}\left[\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}^*\right) \mathbf{Z}_r^*\left(\mathbf{u},\boldsymbol{\theta}^*\right)\right] \\
&\times \left[\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) K\left(\frac{\mathbf{X}_{r-1}-\mathbf{X}_{t-1}}{h}\right)\right. \\
&\quad \left. - \int \frac{a(\mathbf{x})}{g(\mathbf{x})} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) K\left(\frac{\mathbf{X}_{r-1}-\mathbf{x}}{h}\right) d\mathbf{x}\right] dW(\mathbf{u}) \\
&= R_{1A} + R_{2A} + R_{3A} + R_{4A}.
\end{aligned} \tag{A.11}
$$

It is easy to see $R_{jA} = o_P(1)$ for $j = 1,2,3$. We just focus on the proof of $R_{4A} = o_P(1)$.

We further decompose R_{4A}. Define

$$
\begin{aligned}
\phi_A\left(\boldsymbol{\xi}_s, \boldsymbol{\xi}_r, \boldsymbol{\xi}_t\right) = &\int \operatorname{Re}\left[\mathbf{Z}_s\left(\mathbf{u},\boldsymbol{\theta}^*\right) \mathbf{Z}_r^*\left(\mathbf{u},\boldsymbol{\theta}^*\right)\right] \\
&\times \left[\frac{a\left(\mathbf{X}_{t-1}\right)}{g^2\left(\mathbf{X}_{t-1}\right)} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) K\left(\frac{\mathbf{X}_{r-1}-\mathbf{X}_{t-1}}{h}\right)\right. \\
&\quad \left. - \int \frac{a(\mathbf{x})}{g(\mathbf{x})} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) K\left(\frac{\mathbf{X}_{r-1}-\mathbf{x}}{h}\right) d\mathbf{x}\right] dW(\mathbf{u}) \\
&+ \int \operatorname{Re}\left[\mathbf{Z}_r\left(\mathbf{u},\boldsymbol{\theta}^*\right) \mathbf{Z}_t^*\left(\mathbf{u},\boldsymbol{\theta}^*\right)\right] \\
&\times \left[\frac{a\left(\mathbf{X}_{s-1}\right)}{g^2\left(\mathbf{X}_{s-1}\right)} K\left(\frac{\mathbf{X}_{r-1}-\mathbf{X}_{s-1}}{h}\right) K\left(\frac{\mathbf{X}_{t-1}-\mathbf{X}_{s-1}}{h}\right)\right. \\
&\quad \left. - \int \frac{a(\mathbf{x})}{g(\mathbf{x})} K\left(\frac{\mathbf{X}_{r-1}-\mathbf{x}}{h}\right) K\left(\frac{\mathbf{X}_{t-1}-\mathbf{x}}{h}\right) d\mathbf{x}\right] dW(\mathbf{u}) \\
&+ \int \operatorname{Re}\left[\mathbf{Z}_t\left(\mathbf{u},\boldsymbol{\theta}^*\right) \mathbf{Z}_s^*\left(\mathbf{u},\boldsymbol{\theta}^*\right)\right] \\
&\times \left[\frac{a\left(\mathbf{X}_{r-1}\right)}{g^2\left(\mathbf{X}_{r-1}\right)} K\left(\frac{\mathbf{X}_{t-1}-\mathbf{X}_{r-1}}{h}\right) K\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{r-1}}{h}\right)\right. \\
&\quad \left. - \int \frac{a(\mathbf{x})}{g(\mathbf{x})} K\left(\frac{\mathbf{X}_{t-1}-\mathbf{x}}{h}\right) K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) d\mathbf{x}\right] dW(\mathbf{u}),
\end{aligned}
$$

$\phi_A(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r) \equiv \int \phi_A(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r,\boldsymbol{\xi}_t)\,dP(\boldsymbol{\xi}_t)$, $\phi_A(\boldsymbol{\xi}_s) \equiv \int \phi_A(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r)\,dP(\boldsymbol{\xi}_r)$, $\phi_A \equiv \int \phi_A(\boldsymbol{\xi}_s)\,dP(\boldsymbol{\xi}_s)$. We can rewrite R_{4A} as

$$\begin{aligned} R_{4A} = {} & \frac{1}{3T^3h^{2d}} \sum_{t\neq s\neq r} [\phi_A(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r,\boldsymbol{\xi}_t) - \phi_A(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r) - \phi_A(\boldsymbol{\xi}_r,\boldsymbol{\xi}_t) \\ & \qquad - \phi_A(\boldsymbol{\xi}_s,\boldsymbol{\xi}_t) + \phi_A(\boldsymbol{\xi}_s) + \phi_A(\boldsymbol{\xi}_r) + \phi_A(\boldsymbol{\xi}_t)] \\ & + \frac{T-2}{T^3h^{2d}} \sum_{s\neq r} [\phi_A(\boldsymbol{\xi}_s,\boldsymbol{\xi}_r) - \phi_A(\boldsymbol{\xi}_s) - \phi_A(\boldsymbol{\xi}_r)] + \frac{T^2-4T+3}{T^3h^{2d}} \sum_{t=2}^{T} \phi(\boldsymbol{\xi}_t) \\ \equiv {} & R_{41A} + R_{42A} + R_{43A}, \quad \text{say,} \end{aligned} \tag{A.12}$$

where we have used the fact that $\phi_A = 0$.

Proofs of $R_{4jA} = o_P(1)$ for $j = 1,2,3$, are quite similar to proofs of Lemmas A.7, A.8, and A.5, respectively, and are omitted. Theorem A.6 follows from the fact that $R_{jA} = o_P(1)$ for $j = 1,2,3,4$. ■

Proof of Theorem A.7. It suffices to show that $T^{-1}h^{-d/2}\tilde{C} = o_P(1)$, $\tilde{D} - D = o_P(1)$, and

$$\begin{aligned} & \iint \left[\left| \frac{1}{Th^d g(\mathbf{x})} \sum_{s=2}^{T} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) \mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}^*) \right|^2 - |m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}^*)|^2 \right] \\ & \quad \times a(\mathbf{x})g(\mathbf{x})\,d\mathbf{x}\,dW(\mathbf{u}) = o_P(1). \end{aligned} \tag{A.13}$$

Because the proofs for $T^{-1}h^{-d/2}\tilde{C} = o_P(1)$ and $\tilde{D} - D = o_P(1)$ are straightforward, we focus on the proof of (A.13). From the Cauchy–Schwarz inequality and the fact that $\iint |m(\mathbf{u},\mathbf{x})|^2 a(\mathbf{x})\, g(\mathbf{x}) dx\, dW(\mathbf{u}) = O_P(1)$, it suffices to show $\iint \left| \frac{1}{Th^d g(\mathbf{x})} \sum_{s=2}^{T} K\left((\mathbf{X}_{s-1}-\mathbf{x})/h\right) \mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}^*) - m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}^*) \right|^2 a(\mathbf{x})\, g(\mathbf{x}) d\mathbf{x}\, dW(\mathbf{u}) \xrightarrow{p} 0$.

We decompose

$$\begin{aligned} & \mathrm{E}\left| \frac{1}{Th^d g(\mathbf{x})} \sum_{s=2}^{T} K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) \mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}^*) - m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}^*) \right|^2 \\ & = \frac{1}{T^2h^{2d}g(\mathbf{x})^2} \mathrm{E}\left| \sum_{s=2}^{T} \left\{ K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) \mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}^*) \right.\right. \\ & \qquad \left.\left. - \mathrm{E}\left[K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) \mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}^*) \right] \right\} \right|^2 \\ & \quad + \left| \frac{1}{h^d g(\mathbf{x})} \mathrm{E}\left[K\left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right) \mathbf{Z}_s(\mathbf{u},\boldsymbol{\theta}^*) \right] - m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}^*) \right|^2 \\ & = O\left(T^{-1}h^{-d}\right) + O\left(h^2\right) = o(1). \end{aligned}$$

Theorem A.7 follows by Markov's inequality and Assumption A.5. ■

Proof of Theorem 3(ii). Under the assumptions of Theorem 3, we have $\iint_S |m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}^*)|^2 a(\mathbf{x})g(\mathbf{x})d\mathbf{x}dW(\mathbf{u}) > 0$. It follows that $\hat{M} = Th^{d/2}(2D)^{-1/2}\iint_S |m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}^*)|^2 a(\mathbf{x})g(\mathbf{x})d\mathbf{x}dW(\mathbf{u}) + Th^{d/2}(2D)^{-1/2}\iint_{S^C} |m(\mathbf{u},\mathbf{x},\boldsymbol{\theta}^*)|^2 a(\mathbf{x})g(\mathbf{x})d\mathbf{x} + Th^{d/2}o_P(1) \xrightarrow{P} \infty$, where S^c is the complement of the set S, and hence $P[\hat{M} > C(T)] \to 1$ for any sequence of constants $\{C(T) = o(Th^{d/2})\}$. Thus $\hat{M}$ has asymptotic unit power at any given significance level, whenever S has a positive measure on support $\mathbb{W}\times\mathbf{G}$. ■

Proof of Theorem 4. The proof is very similar to the proof of Theorem 1, with $\mathbf{Z}_t^{(\nu)}(\hat{\boldsymbol{\theta}})$, $\mathbf{Z}_t^{(\nu)}(\boldsymbol{\theta}_0)$, $\hat{m}^{(\nu)}(\mathbf{0},\mathbf{x},\hat{\boldsymbol{\theta}})$, $\hat{C}^{(\nu)}$, $\hat{D}^{(\nu)}$ replacing $\mathbf{Z}_t(\hat{\boldsymbol{\theta}})$, $\mathbf{Z}_t(\boldsymbol{\theta}_0)$, $\hat{m}(\mathbf{u},\mathbf{x},\hat{\boldsymbol{\theta}})$, $\hat{C}$, $\hat{D}$, respectively. ■

To prove Theorems 5 and 6, we first provide conditions that are simple modified versions of Assumptions A.1–A.5 and A.7.

Assumption B.1. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space.

(a) The stochastic time series vector process $\mathbf{X}_t = (\mathbf{X}_{1,t}', \mathbf{X}_{2,t}')'$ is a $d\times 1$ strictly stationary continuous-time Markov process, where $\mathbf{X}_{1,t} \subset \mathbb{R}^{d_1}$ denotes the observable state variables, $\mathbf{X}_{2,t} \subset \mathbb{R}^{d_2}$ denotes the unobservable state variables, and $d_1 + d_2 = d$. Here $\mathbf{Y}_t = (\mathbf{X}_{1,t}', \mathbf{X}_{1,t-1}', \ldots, \mathbf{X}_{1,t-l+1}')'$ has the marginal density $f(\mathbf{y}) > 0$ for all $y \in \mathbf{F}$, where $\mathbf{F}$ is a compact set of $\mathbb{R}^{ld_1}$. Also, the joint density of $(\mathbf{Y}_1, \mathbf{Y}_j)$ is bounded by some constant independent of $j > 1$.

(b) A discrete sample $\{\mathbf{X}_{1,t}\}_{t=\Delta}^{T\Delta}$, where $\Delta \equiv 1$ is the sampling interval, is observed at equally spaced discrete times, and $\{\mathbf{X}_t\}_{t=\Delta}^{T\Delta}$ is a β-mixing process with mixing coefficients satisfying $\sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} < C$ for some $0 < \delta < 1$.

Assumption B.2. Let $\phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$ be the CCF of $\mathbf{X}_{1,t}$ given $\mathcal{I}_{1,t-1} \equiv \{\mathbf{X}_{1,t-1}, \mathbf{X}_{1,t-2}, \ldots\}$

(a) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, each $\mathbf{u}_1 \in \mathbb{R}^{d_1}$, and each t, $\phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$ is measurable with respect to $\mathcal{I}_{1,t-1}$.

(b) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, each $\mathbf{u}_1 \in \mathbb{R}^{d_1}$, and each t, $\phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with probability one; and

(c) $\sup_{\mathbf{u}_1\in\mathbb{R}^{d_1}} E\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\frac{\partial}{\partial\boldsymbol{\theta}}\phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})||^2 \le C$ and $\sup_{\mathbf{u}_1\in\mathbb{R}^{d_1}} E\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\phi(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta})|| \le C$.

Assumption B.3. $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1)$, where $\boldsymbol{\theta}^* = p\lim_{T\to\infty}\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ under $\mathbb{H}_0$.

Assumption B.4. The function $K : \mathbb{R}^{ld_1} \to \mathbb{R}^+$ is a product kernel of some univariate kernel k, i.e., $K(\mathbf{u}) = \prod_{j=1}^{ld_1} k(u_j)$, where $k(\cdot) : \mathbb{R} \to \mathbb{R}^+$, which satisfies the Lipschitz condition and is a symmetric, bounded, and twice continuously differentiable function with $\int_{-\infty}^{\infty} k(u)du = 1$, $\int_{-\infty}^{\infty} uk(u)du = 0$, and $\int_{-\infty}^{\infty} u^2k(u)du < \infty$.

Assumption B.5.

(a) $W : \mathbb{R}^{ld_1} \to \mathbb{R}^+$ is a nondecreasing right-continuous weighting function that weighs sets symmetric about the origin equally, with $\int_{\mathbb{R}^{ld_1}} \|\mathbf{u}\|^4 \, dW(\mathbf{u}) < \infty$.

(b) $a : \mathbf{F} \to \mathbb{R}^+$ is a bounded weighting function, where $\mathbf{F} \in \mathbb{R}^{ld_1}$ is the compact support given in Assumption B.1.

Assumption B.6. $\hat{\phi}\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)$ is an estimator for $\phi\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)$ such that $\sqrt{T}\sup_{\mathbf{u}_1 \in \mathbb{R}^{d_1}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{\phi}\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right) - \phi\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)| = O_P(1)$.

Assumption B.7. $\hat{\boldsymbol{\theta}}$ is a parameter estimator such that $p\lim_{T\to\infty} \boldsymbol{\theta} \hat{=} \theta^*$.

Assumption B.8. $\hat{\phi}\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)$ is an estimator for $\phi\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)$ such that $\sup_{\mathbf{u}_1 \in \mathbb{R}^{d_1}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{\phi}\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right) - \phi\left(\mathbf{u}_1, t|\mathcal{I}_{1,t-1}, \boldsymbol{\theta}\right)| = o_P(1)$.

THEOREM A.8. *Under the conditions of Theorem 5, $\hat{M}_u - \bar{M}_u \xrightarrow{p} 0$, where $\bar{M}_u$ is defined in the same way as $\hat{M}_u$ with $\hat{\varphi}\left(\mathbf{u}_1, \mathbf{Y}_s|\mathcal{I}_{s-1}, \hat{\boldsymbol{\theta}}\right)$ replaced by $\phi\left(\mathbf{u}_1, \mathbf{Y}_s|\mathcal{I}_{s-1}, \hat{\boldsymbol{\theta}}\right)$.*

Proof of Theorem A.8. Let $\bar{m}(\mathbf{u}_1, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}})$ be defined in the same way as $\hat{m}(\mathbf{u}_1, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}})$ in (3.3), with $\hat{\phi}\left(\mathbf{u}_1, \mathbf{Y}_s|\mathcal{I}_{s-1}, \hat{\boldsymbol{\theta}}\right)$ replaced by $\phi\left(\mathbf{u}_1, \mathbf{Y}_s|\mathcal{I}_{s-1}, \hat{\boldsymbol{\theta}}\right)$. To show $\hat{M} - \tilde{M} \xrightarrow{p} 0$, it suffices to show $\hat{C}_u - \bar{C}_u = o_P(1)$, $\hat{D}_u - \bar{D}_u = o_P(1)$, and

$$\frac{h^{d/2}}{\sqrt{2\hat{D}_u}} \sum_{t=2}^{T} \int \left[\left|\hat{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}})\right|^2 - \left|\bar{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}})\right|^2\right] \times a\left(\mathbf{Y}_{t-1}\right) dW(\mathbf{u}_1) = o_P(1). \tag{A.14}$$

We focus on the proof of (A.14); the proofs for $\hat{C}_u - \bar{C}_u = o_P(1)$ and $\hat{D}_u - \bar{D}_u = o_P(1)$ are tedious but straightforward.

To show (A.14), we first decompose

$$h^{d/2} \sum_{t=2}^{T} \int \left[\left|\hat{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}})\right|^2 - \left|\bar{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}})\right|^2\right] a\left(\mathbf{Y}_{t-1}\right) dW(\mathbf{u}_1) = \bar{A}_1 + 2\mathrm{Re}(\bar{A}_2), \quad \text{where} \tag{A.15}$$

$$\bar{A}_1 = h^{d/2} \sum_{t=2}^{T} \int \left|\hat{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}}) - \bar{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}})\right|^2 a\left(\mathbf{Y}_{t-1}\right) dW(\mathbf{u}_1),$$

$$\bar{A}_2 = h^{d/2} \sum_{t=2}^{T} \int \left[\hat{m}(\mathbf{u}_1, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}}) - \bar{m}(\mathbf{u}_1, \mathbf{X}_{t-1}, \hat{\boldsymbol{\theta}})\right] \times \bar{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}})^* a\left(\mathbf{Y}_{t-1}\right) dW(\mathbf{u}_1),$$

$\mathrm{Re}(\bar{A}_2)$ is the real part of $\bar{A}_2$, and $\bar{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}})^*$ is the complex conjugate of $\bar{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}})$. Then, (A.14) follows from Propositions A.5 and A.6 as $T \to \infty$.

PROPOSITION A.5. *Under the conditions of Theorem 1,* $\bar{A}_1 \xrightarrow{p} 0$.

PROPOSITION A.6. *Under the conditions of Theorem 1,* $\bar{A}_2 \xrightarrow{p} 0$.

Proof of Proposition A.5. First we have

$$\begin{aligned}
&\hat{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}}) - \bar{m}(\mathbf{u}_1, \mathbf{Y}_{t-1}, \hat{\boldsymbol{\theta}}) \\
&\quad = \frac{1}{Th^d g\left(\mathbf{X}_{t-1}\right)} \sum_{s=2}^{T} K\left(\frac{\mathbf{Y}_{s-1} - \mathbf{Y}_{t-1}}{h}\right) \\
&\qquad \times \left[\mathbf{Z}_{1,s}\left(\mathbf{u}_1, \hat{\boldsymbol{\theta}}\right) - \bar{Z}_{1,s}\left(\mathbf{u}_1, \hat{\boldsymbol{\theta}}\right)\right][1 + O_P(\alpha_T)] \\
&\quad = \frac{1}{Th^d g\left(\mathbf{Y}_{t-1}\right)} \sum_{s=2}^{T} K\left(\frac{\mathbf{Y}_{s-1} - \mathbf{Y}_{t-1}}{h}\right) \\
&\qquad \times \left[\phi\left(\mathbf{u}_1, \mathbf{Y}_s | \mathcal{I}_{s-1}, \hat{\boldsymbol{\theta}}\right) - \hat{\phi}\left(\mathbf{u}_1, \mathbf{Y}_s | \mathcal{I}_{s-1}, \hat{\boldsymbol{\theta}}\right)\right][1 + O_P(\alpha_T)] \\
&\quad = \bar{A}_{11}(\mathbf{u}_1, \mathbf{Y}_{t-1})[1 + O_P(\alpha_T)], \text{ say.}
\end{aligned}$$

Note that we have used Proposition 6.2 in Fan and Yao (2003, p. 236). Proposition A.5 follows from Lemma A.13.

LEMMA A.13. $h^{d/2} \sum_{t=2}^{T} \int \left|\bar{A}_{11}(\mathbf{u}_1, \mathbf{Y}_{t-1})\right|^2 a\left(\mathbf{Y}_{t-1}\right) dW(\mathbf{u}_1) = o_P(1)$.

Proof of Lemma A.13. We decompose

$$\begin{aligned}
&h^{d/2} \sum_{t=2}^{T} \int \left|\bar{A}_{11}(\mathbf{u}_1, \mathbf{Y}_{t-1})\right|^2 a\left(\mathbf{Y}_{t-1}\right) dW(\mathbf{u}_1) \\
&\quad \le \sum_{t=2}^{T} \int \frac{a\left(\mathbf{Y}_{t-1}\right)}{T^2 h^{3d/2} g^2\left(\mathbf{Y}_{t-1}\right)} K(\mathbf{0})^2 \\
&\qquad \times \left|\phi\left(\mathbf{u}_1, t | \mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}\right) - \hat{\phi}\left(\mathbf{u}_1, t | \mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}\right)\right|^2 dW(\mathbf{u}_1) \\
&\qquad + C \sum_{s>t} \int \frac{a\left(\mathbf{Y}_{t-1}\right)}{T^2 h^{3d/2} g^2\left(\mathbf{Y}_{t-1}\right)} K(\mathbf{0}) K\left(\frac{\mathbf{Y}_{s-1} - \mathbf{Y}_{t-1}}{h}\right) \\
&\qquad \times \left|\left[\phi\left(\mathbf{u}_1, t | \mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}\right) - \hat{\phi}\left(\mathbf{u}_1, t | \mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}\right)\right]\right. \\
&\qquad \times \left.\left[\phi\left(\mathbf{u}_1, s | \mathcal{I}_{s-1}, \hat{\boldsymbol{\theta}}\right) - \hat{\phi}\left(\mathbf{u}_1, s | \mathcal{I}_{s-1}, \hat{\boldsymbol{\theta}}\right)\right]^*\right| dW(\mathbf{u}_1) \\
&\qquad + C \sum_{s>t} \int \frac{a\left(\mathbf{Y}_{t-1}\right)}{T^2 h^{3d/2} g^2\left(\mathbf{Y}_{t-1}\right)} K\left(\frac{\mathbf{Y}_{s-1} - \mathbf{Y}_{t-1}}{h}\right)^2 \\
&\qquad \times \left|\phi\left(\mathbf{u}_1, t | \mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}\right) - \hat{\phi}\left(\mathbf{u}_1, t | \mathcal{I}_{t-1}, \hat{\boldsymbol{\theta}}\right)\right|^2 dW(\mathbf{u}_1)
\end{aligned}$$

$$
\begin{aligned}
&+C\sum_{s'>s>t}\int \frac{a\left(\mathbf{Y}_{t-1}\right)}{T^2h^{3d/2}g^2\left(\mathbf{Y}_{t-1}\right)}K\left(\frac{\mathbf{Y}_{s-1}-\mathbf{Y}_{t-1}}{h}\right)K\left(\frac{\mathbf{Y}_{s'-1}-\mathbf{Y}_{t-1}}{h}\right)\\
&\times\left|\left[\phi\left(\mathbf{u}_1,t|\mathcal{I}_{t-1},\hat{\boldsymbol{\theta}}\right)-\hat{\phi}\left(\mathbf{u}_1,t|\mathcal{I}_{t-1},\hat{\boldsymbol{\theta}}\right)\right]\right.\\
&\times\left.\left[\phi\left(\mathbf{u}_1,s|\mathcal{I}_{s-1},\hat{\boldsymbol{\theta}}\right)-\hat{\phi}\left(\mathbf{u}_1,s|\mathcal{I}_{s-1},\hat{\boldsymbol{\theta}}\right)\right]^*\right|dW(\mathbf{u}_1)\\
&=O_P(T^{-2}h^{-3d/2}+T^{-1}h^{-d/2}+T^{-1}h^{-d/2}+h^{d/2})=o_P(1),
\end{aligned}
$$

where we have used Assumption B.8, Markov's inequality, and change of variable. ■

Proof of Proposition A.6. We have

$$
\begin{aligned}
&h^{d/2}\sum_{t=2}^{T}\int\left|\left[\hat{m}(\mathbf{u}_1,\mathbf{Y}_{t-1},\hat{\boldsymbol{\theta}})-\bar{m}(\mathbf{u}_1,\mathbf{Y}_{t-1},\hat{\boldsymbol{\theta}})\right]\bar{m}(\mathbf{u}_1,\mathbf{Y}_{t-1},\hat{\boldsymbol{\theta}})^*\right|a\left(\mathbf{Y}_{t-1}\right)dW(\mathbf{u}_1)\\
&\le\sum_{t=2}^{T}\int\frac{a\left(\mathbf{Y}_{t-1}\right)}{T^3h^{2d}g^2\left(\mathbf{Y}_{t-1}\right)}\left|\sum_{s=2}^{T}K\left(\frac{\mathbf{Y}_{s-1}-\mathbf{Y}_{t-1}}{h}\right)\right.\\
&\quad\times\left.\left[\mathbf{Z}_{1,s}\left(\mathbf{u}_1,\hat{\boldsymbol{\theta}}\right)-\bar{\mathbf{Z}}_{1,s}\left(\mathbf{u}_1,\hat{\boldsymbol{\theta}}\right)\right]\right|\\
&\quad\times\left|\sum_{s=2}^{T}K\left(\frac{\mathbf{Y}_{s-1}-\mathbf{Y}_{t-1}}{h}\right)\bar{\mathbf{Z}}_{1,s}^*\left(\mathbf{u}_1,\hat{\boldsymbol{\theta}}\right)\right|dW\left(\mathbf{u}_1\right)[1+o_P(\alpha_T)]\\
&\le\int\frac{a(\mathbf{y})}{Th^{3d/2}g(\mathbf{y})}\left|\sum_{s=2}^{T}K\left(\frac{\mathbf{Y}_{s-1}-\mathbf{Y}_{t-1}}{h}\right)\right.\\
&\quad\times\left.\left[\phi\left(\mathbf{u}_1,s|\mathcal{I}_{s-1},\hat{\boldsymbol{\theta}}\right)-\hat{\phi}\left(\mathbf{u}_1,s|\mathcal{I}_{s-1},\hat{\boldsymbol{\theta}}\right)\right]\right|\\
&\quad\times\left|\sum_{s=2}^{T}K\left(\frac{\mathbf{Y}_{s-1}-\mathbf{Y}_{t-1}}{h}\right)\bar{\mathbf{Z}}_{1,s}^*\left(\mathbf{u}_1,\hat{\boldsymbol{\theta}}\right)\right|dW\left(\mathbf{u}_1\right)[1+o_P(\alpha_T)]\\
&=O_P\left(T^{-1/2}h^{-d}\right)=o_P(1).
\end{aligned}
$$

■

Proof of Theorem 5. The proof is very similar to the proof of Theorem 1, with $\mathbf{Z}_{1,t}(\hat{\boldsymbol{\theta}})$, $\mathbf{Z}_{1,t}(\boldsymbol{\theta}_0)$, $\hat{m}(\mathbf{u}_1,\mathbf{y},\hat{\boldsymbol{\theta}})$, $\hat{C}_u$, $\hat{D}_u$ replacing $\mathbf{Z}_t(\hat{\boldsymbol{\theta}})$, $\mathbf{Z}_t(\boldsymbol{\theta}_0)$, $\hat{m}(\mathbf{u},\mathbf{x},\hat{\boldsymbol{\theta}})$, $\hat{C}$, $\hat{D}$, respectively. ■

Proof of Theorem 6. The proof is very similar to the proof of Theorem 3, with $\mathbf{Z}_{1,t}(\hat{\boldsymbol{\theta}})$, $\mathbf{Z}_{1,t}(\boldsymbol{\theta}_0)$, $\hat{m}(\mathbf{u}_1,\mathbf{y},\hat{\boldsymbol{\theta}})$, $\hat{C}_u$, $\hat{D}_u$ replacing $\mathbf{Z}_t(\hat{\boldsymbol{\theta}})$, $\mathbf{Z}_t(\boldsymbol{\theta}_0)$, $\hat{m}(\mathbf{u},\mathbf{x},\hat{\boldsymbol{\theta}})$, $\hat{C}$, $\hat{D}$, respectively. ■