Econometric Theory, **28**, 2012, 130–178.
doi:10.1017/S0266466611000065

TESTING FOR THE MARKOV PROPERTY IN TIME SERIES

Bin Chen
University of Rochester

Yongmiao Hong
Cornell University and Xiamen University

The Markov property is a fundamental property in time series analysis and is often assumed in economic and financial modeling. We develop a new test for the Markov property using the conditional characteristic function embedded in a frequency domain approach, which checks the implication of the Markov property in every conditional moment (if it exists) and over many lags. The proposed test is applicable to both univariate and multivariate time series with discrete or continuous distributions. Simulation studies show that with the use of a smoothed nonparametric transition density-based bootstrap procedure, the proposed test has reasonable sizes and all-around power against several popular non-Markov alternatives in finite samples. We apply the test to a number of financial time series and find some evidence against the Markov property.

1. INTRODUCTION

The Markov property is a fundamental property in time series analysis and is often a maintained assumption in economic and financial modeling. Testing for the validity of the Markov property has important implications in economics, finance, as well as time series analysis. In economics, for example, Markov decision processes (MDP), which are based on the Markov assumption, provide a general framework for modeling sequential decision making under uncertainty (see Rust, 1994, and Ljungqvist and Sargent, 2000, for excellent surveys) and have been extensively used in economics, finance, and marketing. Applications of MDP include investment under uncertainty (Lucas and Prescott, 1971; Sargent, 1987), asset pricing (Lucas, 1978; Hall, 1978; Mehra and Prescott, 1985), economic growth (Uzawa, 1965; Romer, 1986, 1990; Lucas, 1988), optimal taxation (Lucas

We thank Pentti Saikkonen (the co-editor), three referees, Frank Diebold, Oliver Linton, James MacKinnon, Katsumi Shimotsu, Kyungchul Song, Liangjun Su, George Tauchen, and seminar participants at Peking University, Queen's University, University of Pennsylvania, the 2008 Xiamen University-Humboldt University Joint Workshop, the 2008 International Symposium on Recent Developments of Time Series Econometrics in Xiamen, the 2008 Symposium on Econometric Theory and Applications (SETA) in Seoul, and the 2008 Far Eastern Econometric Society Meeting in Singapore for their constructive comments on the previous versions of this paper. Any remaining errors are solely ours. Bin Chen thanks the Department of Economics, University of Rochester, for financial support. Yongmiao Hong thanks the outstanding overseas youth fund of the National Science Foundation of China for its support.
Address correspondence to Bin Chen, Department of Economics, University of Rochester, Rochester, NY 14620, USA; e-mail: bchen8@mail.rochester.edu.

and Stokey, 1983; Zhu, 1992), industrial organization (Ericson and Pakes, 1995; Weintraub, Benkard, and Van Roy, 2008), and equilibrium business cycles (Kydland and Prescott, 1982). In the MDP framework, an optimal decision rule can be found within the subclass of nonrandomized Markovian strategies, where a strategy depends on the past history of the process only via the current state. Obviously, the optimal decision rule may be suboptimal if the foundational assumption of the Markov property is violated. Recently non-Markov decision processes (NMDP) have attracted increasing attention in the literature (e.g., Mizutani and Dreyfus, 2004; Aviv and Pazgal, 2005). The non-Markov nature can arise in many ways. The most direct extension of MDP to NMDP is to deprive the decision maker of perfect information on the state of the environment.

In finance the Markov property is one of the most popular assumptions in most continuous-time modeling. It is well known that stochastic integrals yield Markov processes. In modeling interest rate term structure, such popular models as Vasicek (1977), Cox, Ingersoll, and Ross (1985), affine term structure models (Duffie and Kan, 1996; Dai and Singleton, 2000), quadratic term structure models (Ahn, Dittmar, and Gallant, 2002), and affine jump diffusion models (Duffie, Pan, and Singleton, 2000) are all Markov processes. They are widely used in pricing and hedging fixed-income or equity derivatives, managing financial risk, and evaluating monetary policy and debt policy. If interest rate processes are not Markov, alternative non-Markov models, such as Heath, Jarrow, and Morton's (1992) model may provide a better characterization of interest rate dynamics. In a discrete-time framework, Duan and Jacobs (2008) find that deviations from the Markovian structure significantly improve the empirical performance of the model for the short-term interest rate. In general, if a process is obtained by discretely sampling a subset of the state variables of a continuous-time process that evolves according to a system of nonlinear stochastic differential equations, it is non-Markov. A leading example is the class of stochastic volatility models (e.g., Anderson and Lund, 1997; Gallant, Hsieh, and Tauchen, 1997).

In the market microstructure literature, one important issue is the price formation mechanism, which determines whether security prices follow a Markov process. Easley and O'Hara (1987) develop a structural model of the effect of asymmetric information on the price-trade size relationship. They show that trade size introduces an adverse selection problem to security trading because informed traders, given their wish to trade, prefer to trade larger amounts at any given price. Hence market makers' pricing strategies will also depend on trade size, and the entire sequence of past trades is informative of the likelihood of an information event and thus price evolution. Consequently, prices typically will not follow a Markov process. Easley and O'Hara (1992) further consider a variant of Easley and O'Hara's (1987) model and delineate the link between the existence of information, the timing of trades, and the stochastic process of security prices. They show that while trade signals the direction of any new information, the lack of trade signals the existence of any new information. The latter effect can be viewed as event uncertainty and suggests that the interval between trades may be informative and hence time per se is not exogenous to the price process. One implication of this model is that either quotes or prices combined with inventory, volume, and clock time are Markov processes. Therefore, rather than using the non-Markov price series alone, it would be preferable to estimate the price process consisting of no trade outcomes, buys, and sells. On the other hand, other models also explain market behavior but reach opposite conclusions on the property of prices. For example, Platen and Rebolledo (1996) and Amaro de Matos and Rosario (2000) propose equilibrium models, which assume that market makers can take advantage of their superior information on trade orders and set different prices. The presence of market makers prevents the direct interaction between demand and supply sides. By specifying the supply and demand processes, these market makers obtain the equilibrium prices, which may be Markov. By testing the Markov property, one can check which models reflect reality more closely.

Our interest in testing the Markov property is also motivated by its wide applications among practitioners. For example, technical analysis has been used widely in financial markets for decades (see, e.g., Edwards and Magee, 1966; Blume, Easley, and O'Hara, 1994; LeBaron, 1999). One important category is priced-based technical strategies, which refer to the forecasts based on past prices, often via moving-average rules. However, if the history of prices does not provide additional information, in the sense that the current prices already impound all information, then price-based technical strategies would not be effective. In other words, if prices adjust immediately to information, past prices would be redundant and current prices are the sufficient statistics for forecasting future prices. This actually corresponds to a fundamental issue: namely, whether prices follow a Markov process.

Finally, in risk management, financial institutions are required to rate assets by their default probability and by their expected loss severity given a default. For this purpose, historical information on the transition of credit exposures is used to estimate various models that describe the probabilistic evolution of credit quality. The simple time-homogeneous Markov model is one of the most popular models (e.g., Jarrow and Turnbull, 1995; Jarrow, Lando, and Turnbull, 1997), specifying the stochastic processes completely by transition probabilities. Under this model, a detailed history of individual assets is not needed. However, whether the Markov specification adequately describes credit rating transitions over time has substantial impact on the effectiveness of credit risk management. In empirical studies, Kavvathas (2001) and Lando and Skϕdeberg (2002) document strong non-Markov behaviors such as dependence on previous rating and waiting-time effects in rating transitions. In contrast, Bangia, Diebold, Kronimus, Schagen, and Schuermann (2002) and Kiefer and Larson (2004) find that first-order Markov ratings dynamics provide a reasonable practical approximation.

Despite innumerable studies rooted in Markov processes, there are few existing tests for the Markov property in the literature. Ait-Sahalia (1997) first proposes a test for whether the interest rate process is Markov by checking the validity of the Chapman-Kolmogorov equation, where the transition density is estimated nonparametrically. The Chapman-Kolmogorov equation is an important characterization of Markov processes and can detect many non-Markov processes with practical importance, but it is only a necessary condition of the Markov property. Feller (1959), Rosenblatt (1960), and Rosenblatt and Slepian (1962) provide examples of stochastic processes that are not Markov but whose first-order transition probabilities nevertheless satisfy the Chapman-Kolmogorov equation. Ait-Sahalia's (1997) test will miss these non-Markov processes.

Amaro de Matos and Fernandes (2007) test whether discretely recorded observations of a continuous-time process are consistent with the Markov property via a smoothed nonparametric density approach. They test the conditional independence of the underlying data generating process (DGP).[1] Because only a fixed lag order in the past information set is checked, the test may easily overlook the violation of conditional independence from higher-order lags. Moreover, the test involves a relatively high-dimensional smoothed nonparametric joint density estimation (see more discussion below).

In this paper we provide a conditional characteristic function (CCF) characterization for the Markov property and use it to construct a nonparametric test for the Markov property. The characteristic function has been widely used in time series analysis and econometrics (e.g., Feuerverger and McDunnough, 1981; Epps and Pulley, 1983; Hong, 1999; Singleton, 2001; Jiang and Knight, 2002; Chacko and Viceira, 2003; and Su and White, 2007). The basic idea of the CCF-characterization for the Markov property is that when and only when a stochastic process is Markov, a generalized residual process associated with the CCF is a martingale difference sequence (MDS). This characterization has never been used in testing the Markov property. We use a nonparametric regression method to estimate the CCF and use a spectral approach to check whether the generalized residuals are explainable by the entire history of the underlying processes. Our approach has several attractive features.

First, we use a novel generalized cross-spectral approach, which embeds the CCF in a spectral framework, thus enjoying the appealing features of spectral analysis. In particular, our approach can examine a growing number of lags as the sample size increases without suffering from the notorious "curse of dimensionality" problem. This improves upon the existing tests, which can only check a fixed number of lags.

Second, as the Fourier transform of the transition density, the CCF can also capture the full dynamics of the underlying process, but it involves a lower dimensional smoothed nonparametric regression than the nonparametric density approaches in the literature.

Third, because we impose regularity conditions directly on the CCF of a discretely observed random sample, our test is applicable to discrete-time processes and continuous-time processes with discretely observed data. It is also applicable to both univariate and multivariate time series processes. Due to the nonparametric nature of the test, it does not need any parametric specification of the underlying process and thus avoids the misspecification problems.

In Section 2 we describe the hypotheses of interest and propose a novel approach to testing the Markov property. We derive the asymptotic distribution of the proposed test statistic in Section 3, and we discuss its asymptotic power property in Section 4. In Section 5 we use Horowitz's (2003) smoothed nonparametric transition-based bootstrap procedure to obtain the finite sample critical values of the test and examine the finite sample performance of the test in comparison with some existing popular tests. We also apply our test to stock prices, interest rates, and foreign exchange rates and document some evidence against the Markov property with all three financial time series. Section 6 concludes. All mathematical proofs are collected in the Appendix. A Gauss code to implement our test is available from the authors upon request. Throughout the paper we 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 AND TEST STATISTICS

Suppose $\{\mathbf{X}_t\}$ is a strictly stationary d-dimensional time series process, where d is a positive integer. It follows a Markov process if the conditional probability distribution of $\mathbf{X}_{t+1}$ given the information set $\mathcal{I}_t = \{\mathbf{X}_t, \mathbf{X}_{t-1}, ...\}$ is the same as the conditional probability distribution of $\mathbf{X}_{t+1}$ given $\mathbf{X}_t$ only. This can be formally expressed as

$$\mathbb{H}_0 : P(\mathbf{X}_{t+1} \leq \mathbf{x}|\mathcal{I}_t) = P(\mathbf{X}_{t+1} \leq \mathbf{x}|\mathbf{X}_t) \quad \text{almost surely (a.s.)} \\ \text{for all } \mathbf{x} \in \mathbb{R}^d \text{ and all } t \geq 1. \tag{2.1}$$

Under $\mathbb{H}_0$, the past information set $\mathcal{I}_{t-1}$ is redundant in the sense that the current state variable or vector $\mathbf{X}_t$ will contain all information about the future behavior of the process that is contained in the current information set $\mathcal{I}_t$. Alternatively, when

$$\mathbb{H}_A : P(\mathbf{X}_{t+1} \leq \mathbf{x}|\mathcal{I}_t) \neq P(\mathbf{X}_{t+1} \leq \mathbf{x}|\mathbf{X}_t) \quad \text{for some } t \geq 1, \tag{2.2}$$

then $\mathbf{X}_t$ is not a Markov process.[2]

Ait-Sahalia (1997) proposes a nonparametric kernel-based test for $\mathbb{H}_0$ by checking the Chapman-Kolmogorov equation,

$$g(\mathbf{X}_{t+1}|\mathbf{X}_{t-1}) = \int_{R^d} g(\mathbf{X}_{t+1}|\mathbf{X}_t = \mathbf{x})\, g(\mathbf{X}_t = \mathbf{x}|\mathbf{X}_{t-1})\, d\mathbf{x} \quad \text{for all } t \geq 1,$$

where $g(\cdot|\cdot)$ is the conditional probability density function estimated by the smoothed nonparametric kernel method. The Chapman-Kolmogorov equation is an important characterization of the Markov property and can detect many non-Markov processes with practical importance. However, there exist non-Markov processes whose first-order transition probabilities satisfy the Chapman-Kolmogorov equation (Feller, 1959; Rosenblatt, 1960; Rosenblatt and Slepian, 1962). Ait-Sahalia's (1997) test is expected to miss these processes.

Amaro de Matos and Fernandes (2007) propose a nonparametric kernel-based test for $\mathbb{H}_0$ by checking the conditional independence between $\mathbf{X}_{t+1}$ and $\mathbf{X}_{t-j}$ given $\mathbf{X}_t$, namely,

$$g\left(\mathbf{X}_{t+1}|\mathbf{X}_t\right) = g\left(\mathbf{X}_{t+1}|\mathbf{X}_t, \mathbf{X}_{t-j}\right) \quad \text{for all } t, j \geq 1,$$

which is implied by $\mathbb{H}_0$. By choosing $j = 1$, Amaro de Matos and Fernandes check

$$g\left(\mathbf{X}_{t+1}, \mathbf{X}_t, \mathbf{X}_{t-1}\right) = g\left(\mathbf{X}_{t+1}|\mathbf{X}_t\right) g\left(\mathbf{X}_t, \mathbf{X}_{t-1}\right) \quad \text{for all } t \geq 1,$$

in their simulation and empirical studies. This approach requires a $3d$-dimensional smoothed nonparametric joint density estimation for $g(\mathbf{X}_{t+1}, \mathbf{X}_t, \mathbf{X}_{t-1})$.

Both of the existing tests essentially check the conditional independence of

$$g(\mathbf{X}_{t+1}|\mathbf{X}_t, \mathbf{X}_{t-1}) = g(\mathbf{X}_{t+1}|\mathbf{X}_t) \quad \text{for all } t \geq 1,$$

which is implied by $\mathbb{H}_0$ in (2.1), but the converse is not true. The most important feature of $\mathbb{H}_0$ is the necessity of checking the entire currently available information $\mathcal{I}_t$. Inevitably there will be information loss if only one lag order is considered. For example, the existing tests may overlook the departure of the Markov property from higher-order lags, say, $\mathbf{X}_{t-2}$. Moreover, their tests may suffer from the curse of dimensionality problem when the dimension d is relatively large, because the nonparametric density estimators $\hat{g}(\mathbf{X}_{t+1}|\mathbf{X}_t, \mathbf{X}_{t-1})$ and $\hat{g}(\mathbf{X}_{t+1}|\mathbf{X}_t)$ involve $3d$ and $2d$ dimensional smoothing, respectively.

We now develop a new test for $\mathbb{H}_0$ using the CCF. As the Fourier transform of the conditional probability density, the CCF can also capture the full dynamics of $\mathbf{X}_{t+1}$. Let $\varphi(u|\mathbf{X}_t)$ be the CCF of $\mathbf{X}_{t+1}$ conditioning on its current state $\mathbf{X}_t$; that is,

$$\varphi(\mathbf{u}|\mathbf{X}_t) = \int_{R^d} e^{i\mathbf{u}'\mathbf{x}} g(\mathbf{x}|\mathbf{X}_t) d\mathbf{x}, \qquad \mathbf{u} \in \mathbb{R}^d, \qquad i = \sqrt{-1}. \tag{2.3}$$

Let $\varphi(u|\mathcal{I}_t)$ be the CCF of $\mathbf{X}_{t+1}$ conditioning on the currently available information $\mathcal{I}_t$, that is,

$$\varphi(\mathbf{u}|\mathcal{I}_t) = \int_{R^d} e^{i\mathbf{u}'\mathbf{x}} g(\mathbf{x}|\mathcal{I}_t) d\mathbf{x}, \qquad \mathbf{u} \in \mathbb{R}^d, \qquad i = \sqrt{-1}.$$

Given the equivalence between the conditional probability density and the CCF, the hypotheses of interest $\mathbb{H}_0$ in (2.1) versus $\mathbb{H}_A$ in (2.2) can be written as

$$\mathbb{H}_0 : \varphi(\mathbf{u}|\mathbf{X}_t) = \varphi(\mathbf{u}|\mathcal{I}_t) \quad \text{a.s. for all } \mathbf{u} \in \mathbb{R}^d \text{ and all } t \geq 1, \tag{2.4}$$

versus the alternative hypothesis

$$\mathbb{H}_A : \varphi(\mathbf{u}|\mathbf{X}_t) \neq \varphi(\mathbf{u}|\mathcal{I}_t) \quad \text{for some } t \geq 1. \tag{2.5}$$

There exist other characterizations of the Markov property. For example, Darsow, Nguyen, and Olsen (1992) and Ibragimov (2007) provide copula-based characterizations of Markov processes. The CCF-based characterization is intuitively appealing and offers much flexibility. To gain insight into this approach, we define a complex-valued process

$$Z_{t+1}(\mathbf{u}) = e^{i\mathbf{u}'\mathbf{X}_{t+1}} - \varphi(\mathbf{u}|\mathbf{X}_t), \qquad \mathbf{u} \in \mathbb{R}^d.$$

Then the Markov property is equivalent to the MDS characterization

$$\mathrm{E}\left[Z_{t+1}(\mathbf{u})|\mathcal{I}_t\right] = 0 \quad \text{for all } \mathbf{u} \in \mathbb{R}^d \text{ and } t \geq 1. \tag{2.6}$$

The process $\{Z_t(\mathbf{u})\}$ may be viewed as a residual of the nonparametric regression

$$e^{i\mathbf{u}'\mathbf{X}_{t+1}} = \mathrm{E}\left(e^{i\mathbf{u}'\mathbf{X}_{t+1}}|\mathbf{X}_t\right) + Z_{t+1}(\mathbf{u}) = \varphi(\mathbf{u}|\mathbf{X}_t) + Z_{t+1}(\mathbf{u}).$$

The MDS characterization in (2.6) has implications on all conditional moments of $\mathbf{X}_t$ when the latter exist. To see this, we consider a Taylor series expansion of (2.6), for the case of $d = 1$,[3] around the origin of $\mathbf{u}$:

$$\mathrm{E}\left[Z_{t+1}(\mathbf{u})|\mathcal{I}_t\right] = \sum_{m=0}^{\infty} \frac{(i\mathbf{u})^m}{m!}\{\mathrm{E}(\mathbf{X}_{t+1}^m|\mathcal{I}_t) - \mathrm{E}(\mathbf{X}_{t+1}^m|\mathbf{X}_t)\} = 0$$
$$\text{for } t \geq 1 \text{ and all } \mathbf{u} \text{ near } 0. \tag{2.7}$$

Thus, checking (2.6) is equivalent to checking whether all conditional moments of $\mathbf{X}_{t+1}$ (if they exist) are Markov. Nevertheless, the use of (2.6) itself does not require any moment conditions of $\mathbf{X}_{t+1}$.

It is not a trivial task to check (2.6). First, the MDS property in (2.6) must hold for all $\mathbf{u} \in \mathbb{R}^d$, not just a finite number of grid points of $\mathbf{u}$. This is an example of the so-called nuisance parameter problem encountered in the literature (e.g., Davies, 1977, 1987; Hansen, 1996). Second, the generalized residual process $Z_{t+1}(\mathbf{u})$ is unknown because the CCF $\varphi(\mathbf{u}|\mathbf{X}_t)$ is unknown, and it has to be estimated nonparametrically to be free of any potential model misspecification. Third, the conditioning information set $\mathcal{I}_t$ in (2.6) has an infinite dimension as $t \to \infty$, so there is a curse of dimensionality difficulty associated with testing the Markov property. Finally, $\{Z_t(\mathbf{u})\}$ may display serial dependence in its higher-order conditional moments. Any test for (2.6) should be robust to time-varying conditional heteroskedasticity and higher-order moments of unknown form in $\{Z_t(\mathbf{u})\}$.

To check the MDS property of $\{Z_t(\mathbf{u})\}$, we extend Hong's (1999) univariate generalized spectrum to a multivariate generalized cross-spectrum.[4] Just as the conventional spectral density is a basic analytic tool for linear time series, the generalized spectrum, which embeds the characteristic function in a spectral framework, is an analytic tool for nonlinear time series. It can capture nonlinear dynamics while maintaining the nice features of spectral analysis, particularly its appealing property to accommodate all lags information. In the present context it can check departures of the Markov property over many lags in a pairwise manner, avoiding the curse of dimensionality difficulty. This is not achievable by the existing tests in the literature. They only check a fixed lag order.

Define the generalized covariance function

$$\Gamma_j(\mathbf{u},\mathbf{v}) = \text{cov}[Z_t(\mathbf{u}), e^{i\mathbf{v}'\mathbf{X}_{t-|j|}}], \qquad \mathbf{u},\mathbf{v} \in \mathbb{R}^d. \tag{2.8}$$

Given that the conventional spectral density is defined as the Fourier transform of the autocovariance function, we can define a generalized cross-spectrum

$$F(\omega,\mathbf{u},\mathbf{v}) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\Gamma_j(\mathbf{u},\mathbf{v})e^{-ij\omega}, \qquad \omega \in [-\pi,\pi], \qquad \mathbf{u},\mathbf{v} \in \mathbb{R}^d, \tag{2.9}$$

which is the Fourier transform of the generalized covariance function $\Gamma_j(\mathbf{u},\mathbf{v})$, where ω is a frequency. This function contains the same information as $\Gamma_j(\mathbf{u},\mathbf{v})$. No moment conditions on $\{\mathbf{X}_t\}$ are required. This is particularly appealing for economic and financial time series. It has been argued that higher moments of financial time series may not exist (e.g., Pagan and Schwert, 1990; Loretan and Phillips, 1994). Moreover, the generalized cross-spectrum can capture cyclical patterns caused by linear and nonlinear cross-dependence, such as volatility clustering and tail clustering of the distribution.

Under $\mathbb{H}_0$ we have $\Gamma_j(\mathbf{u},\mathbf{v}) = 0$ for all $\mathbf{u},\mathbf{v} \in \mathbb{R}^d$ and all $j \neq 0$. Consequently, the generalized cross-spectrum $F(\omega,\mathbf{u},\mathbf{v})$ becomes a "flat" spectrum as a function of frequency ω:

$$F(\omega,\mathbf{u},\mathbf{v}) = F_0(\omega,\mathbf{u},\mathbf{v}) \equiv \frac{1}{2\pi}\Gamma_0(\mathbf{u},\mathbf{v}), \qquad \omega \in [-\pi,\pi], \qquad \mathbf{u},\mathbf{v} \in \mathbb{R}^d. \tag{2.10}$$

Thus, we can test $\mathbb{H}_0$ by checking whether a consistent estimator for $F(\omega,\mathbf{u},\mathbf{v})$ is flat with respect to frequency ω. Any significant deviation from a flat generalized cross-spectrum is evidence of the violation of the Markov property.

The hypothesis of $\mathrm{E}[Z_t(\mathbf{u})|\mathcal{I}_{t-1}] = 0$ for all $\mathbf{u} \in \mathbb{R}^d$ is different from the hypothesis of $\Gamma_j(\mathbf{u},\mathbf{v}) = 0$ for all $\mathbf{u},\mathbf{v} \in \mathbb{R}^d$ and all $j \neq 0$. The former implies the latter but not vice versa. This gap is the price we have to pay for dealing with the difficulty of the curse of dimensionality. From a theoretical point of view, the pairwise approach will miss dependent processes that are pairwise independent. However, such processes apparently do not appear in most empirical applications in economics and finance.

It is rather difficult to formally characterize the gap between $\mathrm{E}\left[Z_t(\mathbf{u})|\mathcal{I}_{t-1}\right] = 0$ for all $\mathbf{u} \in \mathbb{R}^d$ and $\Gamma_j(\mathbf{u},\mathbf{v}) = 0$ for all $\mathbf{u},\mathbf{v} \in \mathbb{R}^d$ and all $j \neq 0$. However, these two hypotheses coincide under some special cases. One example is when $\{\mathbf{X}_t\}$ follows an additive process: $\mathbf{X}_t = \alpha_0 + \sum_{j=1}^{\infty} g(\mathbf{X}_{t-j}) + \varepsilon_t$, where $g(\cdot)$ is not a zero function at least for some lag $j > 0$. Additive time series processes have attracted considerable interest in the nonparametric literature (e.g., Masry and Tjøstheim, 1997; Kim and Linton, 2003).

To reduce the gap between $\mathrm{E}[Z_t(\mathbf{u})|\mathcal{I}_{t-1}] = 0$ for all $\mathbf{u} \in \mathbb{R}^d$ and $\Gamma_j(\mathbf{u},\mathbf{v}) = 0$ for all $\mathbf{u},\mathbf{v} \in \mathbb{R}^d$ and all $j \neq 0$, we can extend $F(\omega,\mathbf{u},\mathbf{v})$ to a generalized bispectrum

$$B(\omega_1,\omega_2,\mathbf{u},\mathbf{v},\tau) = \frac{1}{(2\pi)^2}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} C_{j,l}(\mathbf{u},\mathbf{v},\tau)\, e^{-ij\omega_1 - il\omega_2},$$
$$\omega_1,\omega_2 \in [-\pi,\pi], \qquad \mathbf{u},\mathbf{v},\tau \in \mathbb{R}^d,$$

where

$$C_{j,l}(\mathbf{u},\mathbf{v},\tau) = Z_t(\mathbf{u})\left[e^{i\mathbf{v}'\mathbf{X}_{t-|j|}} - \hat{\varphi}(\mathbf{v})\right]\left[e^{i\tau'\mathbf{X}_{t-|l|}} - \hat{\varphi}(\tau)\right], \qquad \mathbf{u},\mathbf{v},\tau \in \mathbb{R}^d$$

is a generalized third-order central cumulant function. This is equivalent to the use of $\mathrm{E}[Z_t(\mathbf{u})|\mathbf{X}_{t-j},\mathbf{X}_{t-l}]$. With $C_{j,l}(\mathbf{u},\mathbf{v},\tau)$, we can detect a larger class of alternatives to $\mathrm{E}[Z_t(\mathbf{u})|\mathcal{I}_{t-1}] = 0$. Note that the nonparametric generalized bispectrum approach can check many pairs of lags (j,l), while still avoiding the curse of dimensionality. Nevertheless, in this paper, we focus on $\Gamma_j(\mathbf{u},\mathbf{v})$ for simplicity.

Suppose now we have a discretely observed sample $\{\mathbf{X}_t\}_{t=1}^T$ of size T, and we consider consistent estimation of $F(\omega,\mathbf{u},\mathbf{v})$ and $F_0(\omega,\mathbf{u},\mathbf{v})$. Because $Z_t(\mathbf{u})$ is not observable, we have to estimate it first. Then we can estimate the generalized covariance $\Gamma_j(\mathbf{u},\mathbf{v})$ by its sample analogue

$$\hat{\Gamma}_j(\mathbf{u},\mathbf{v}) = \frac{1}{T-|j|}\sum_{t=|j|+1}^{T} \hat{Z}_t(\mathbf{u})\left[e^{i\mathbf{v}'\mathbf{X}_{t-|j|}} - \hat{\varphi}(\mathbf{v})\right], \qquad \mathbf{u},\mathbf{v} \in \mathbb{R}^d, \qquad \mathbf{(2.11)}$$

where the estimated generalized residual

$$\hat{Z}_t(\mathbf{u}) = e^{i\mathbf{u}'\mathbf{X}_t} - \hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1}),$$

$\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$ is a consistent estimator for $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$, and $\hat{\varphi}(\mathbf{v}) = T^{-1}\sum_{t=1}^{T} e^{i\mathbf{v}'\mathbf{X}_t}$ is the empirical characteristic function of $\mathbf{X}_t$. We do not parameterize $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$, which would suffer from potential model misspecification. We use nonparametric regression to estimate $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$. Various nonparametric regression methods could be used here. For concreteness, we use local polynomial regression.

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

Following Masry (1996a, 1996b), we introduce the notations

$$\mathbf{j} = (j_1, ..., j_d), \qquad \mathbf{j}! = j_1! \times ... \times j_d!, \qquad |\mathbf{j}| = \sum_{l=1}^{d} j_l,$$

$$\mathbf{x}^{\mathbf{j}} = x_1^{j_1} \times ... \times x_d^{j_d},$$

$$\sum_{0 \le |\mathbf{j}| \le r} = \sum_{l=0}^{r} \sum_{j_1=0}^{l} \cdots \sum_{\substack{j_d=0 \\ j_1+...+j_d=l}}^{l}.$$

We consider the multivariate local weighted least squares problem

$$\min_{\beta \in \mathbb{R}^N} \sum_{t=2}^{T} \left| e^{i\mathbf{u}'\mathbf{X}_t} - \sum_{0 \le |\mathbf{j}| \le r} \beta_{\mathbf{j}}' (\mathbf{X}_{t-1} - \mathbf{x})^{\mathbf{j}} \right|^2 \mathbf{K}_h (\mathbf{x} - \mathbf{X}_{t-1}), \, \mathbf{x} \in \mathbb{R}^d, \qquad \textbf{(2.12)}$$

where $\beta = \left(\beta_0, \beta_1', .., \beta_r'\right)'$ is an $N \times 1$ parameter vector, $N = \sum_{l=0}^{r} N_l$, $N_l = \frac{(l+d-1)!}{(d-1)!l!}$, $\mathbf{K}_h(\mathbf{x}) = h^{-d}\mathbf{K}(\mathbf{x}/h)$, $\mathbf{K} : \mathbb{R}^d \to \mathbb{R}$ is a kernel function, h is a bandwidth, and r is an odd integer. When r =1, (2.12) boils down to a local linear regression. An example of $\mathbf{K}(\cdot)$ is a prespecified symmetric probability density function. We obtain the following solution to (2.12):

$$\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} \hat{\boldsymbol{\beta}}_0 (\mathbf{x}, \mathbf{u}) \\ \vdots \\ \hat{\boldsymbol{\beta}}_r (\mathbf{x}, \mathbf{u}) \end{bmatrix} = S_T^{-1}(\mathbf{x}) \Gamma(\mathbf{x}, \mathbf{u}), \qquad \mathbf{x} \in \mathbb{R}^d, \qquad \textbf{(2.13)}$$

where $S_T(\mathbf{x})$ is an $N \times N$ matrix

$$\mathbf{S}_T(\mathbf{x}) = \begin{bmatrix} S_{0,0} & S_{0,1} & \cdots & S_{0,r} \\ S_{1,0} & S_{1,1} & \cdots & S_{1,r} \\ \vdots & & & \vdots \\ S_{r,0} & S_{r,1} & \cdots & S_{r,r} \end{bmatrix},$$

$S_{|\mathbf{j}|,|\mathbf{l}|}$ is an $N_{|\mathbf{j}|} \times N_{|\mathbf{l}|}$ matrix with its (m,n)th element $\left(S_{|\mathbf{j}|,|\mathbf{l}|}\right)_{m,n} = s_{g_{|\mathbf{j}|}(m)+g_{|\mathbf{l}|}(n)}$,

$$s_{\mathbf{j}}(\mathbf{x}) = \frac{1}{T-1} \sum_{t=2}^{T} \left(\frac{\mathbf{X}_{t-1} - \mathbf{x}}{h} \right)^{\mathbf{j}} \mathbf{K}_h(\mathbf{X}_{t-1} - \mathbf{x}),$$

and g_l^{-1} denotes the one-to-one map that arranges those N_l d-tuples as a sequence in a lexicographical order.[5] And $\Gamma(\mathbf{x}, \mathbf{u})$ is an $N \times 1$ vector

$$\Gamma(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} \Gamma_0 \\ \Gamma_1 \\ \vdots \\ \Gamma_r \end{bmatrix},$$

$\Gamma_{|\mathbf{j}|}$ is of dimension $N_{|\mathbf{j}|} \times 1$, with its lth element $\left(\Gamma_{|\mathbf{j}|}\right)_l = \tau_{g_{|\mathbf{j}|}(l)}$, and

$$\tau_{\mathbf{j}}(\mathbf{x}) = \sum_{t=2}^{T} e^{i\mathbf{u}'\mathbf{X}_t} \left(\frac{\mathbf{X}_{t-1} - \mathbf{x}}{h}\right)^{\mathbf{j}} \mathbf{K}_h(\mathbf{X}_{t-1} - \mathbf{x}).$$

Note that $\hat{\beta}$ depends on the location $\mathbf{x}$ and parameter $\mathbf{u}$, but for notational simplicity, we have suppressed its dependence on $\mathbf{x}$ and $\mathbf{u}$.

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

$$\hat{\varphi}(\mathbf{u}|\mathbf{x}) = \sum_{t=2}^{T} \hat{W}\left(\frac{\mathbf{X}_{t-1} - \mathbf{x}}{h}\right) e^{i\mathbf{u}'\mathbf{X}_t},$$

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

$$\hat{W}(\mathbf{z}) \equiv (T-1)^{-1} \mathbf{e}_1' \mathbf{S}_T^{-1} \Theta(\mathbf{z}) \mathbf{K}(\mathbf{z}) / h^d,$$

$\mathbf{e}_1 = (1, 0, ..., 0)'$ is an $N \times 1$ unit vector, $\Theta(\mathbf{z})$ is an $N \times 1$ vector

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

$\Theta_{|\mathbf{j}|}(\mathbf{z})$ is of dimension $N_{|\mathbf{j}|} \times 1$, with its lth element $\left[\Theta_{|\mathbf{j}|}(\mathbf{z})\right]_l = (\mathbf{z})^{g_{|\mathbf{j}|}(l)}$, and $\mathbf{z}$ is a $d \times 1$ vector. The regression estimator $\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$ only involves a d-dimensional smoothing, thus enjoying some advantages over the existing nonparametric density approaches, which involve a $2d$ or $3d$ dimensional smoothing.

With the sample generalized covariance function $\hat{\Gamma}_j(\mathbf{u}, \mathbf{v})$, we can construct a consistent estimator for the flat generalized spectrum $F_0(\omega, \mathbf{u}, \mathbf{v})$,

$$\hat{F}_0(\omega, \mathbf{u}, \mathbf{v}) = \frac{1}{2\pi} \hat{\Gamma}_0(\mathbf{u}, \mathbf{v}), \qquad \omega \in [-\pi, \pi], \qquad \mathbf{u}, \mathbf{v} \in \mathbb{R}^d.$$

Consistent estimation for $F(\omega, \mathbf{u}, \mathbf{v})$ is more challenging. We use a nonparametric smoothed kernel estimator for $F(\omega, \mathbf{u}, \mathbf{v})$:

$$\hat{F}(\omega, \mathbf{u}, \mathbf{v}) = \frac{1}{2\pi} \sum_{j=1-T}^{T-1} (1 - |j|/T)^{1/2} k(j/p) \hat{\Gamma}_j(\mathbf{u}, \mathbf{v}) e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \qquad \mathbf{u}, \mathbf{v} \in \mathbb{R}^d, \tag{2.14}$$

where $p = p(T) \to \infty$ is a lag order, and $k : \mathbb{R} \to [-1, 1]$ is a kernel function that assigns weights to various lag orders. Note that $k(\cdot)$ here is different from the kernel $\mathbf{K}(\cdot)$ in (2.12). Most commonly used kernels discount higher-order lags. Examples of commonly used $k(\cdot)$ include the Bartlett kernel

$$k(z) = \begin{cases} 1 - |z|, & |z| \le 1, \\ 0, & \text{otherwise}, \end{cases} \quad \textbf{(2.15)}$$

the Parzen kernel

$$k(z) = \begin{cases} 1 - 6z^2 + 6|z|^3, & |z| \le 0.5, \\ 2(1-|z|)^3, & 0.5 < |z| \le 1, \\ 0, & \text{otherwise}. \end{cases} \quad \textbf{(2.16)}$$

and the quadratic-spectral kernel

$$k(z) = \frac{3}{(\pi z)^2}\left[\frac{\sin(\pi z)}{\pi z} - \cos(\pi z)\right], \qquad \in \mathbb{R}. \quad \textbf{(2.17)}$$

In (2.14) the factor $(1-|j|/T)^{1/2}$ is a finite-sample correction. It could be replaced by unity. Under certain regularity conditions, $\hat{F}(\omega, \mathbf{u}, \mathbf{v})$ and $\hat{F}_0(\omega, \mathbf{u}, \mathbf{v})$ are consistent for $F(\omega, \mathbf{u}, \mathbf{v})$ and $F_0(\omega, \mathbf{u}, \mathbf{v})$, respectively. The estimators $\hat{F}(\omega, \mathbf{u}, \mathbf{v})$ and $\hat{F}_0(\omega, \mathbf{u}, \mathbf{v})$ converge to the same limit under $\mathbb{H}_0$ and generally converge to different limits under $\mathbb{H}_A$. Thus any significant divergence between them is evidence of the violation of the Markov property.

We can measure the distance between $\hat{F}(\omega, \mathbf{u}, \mathbf{v})$ and $\hat{F}_0(\omega, \mathbf{u}, \mathbf{v})$ by the quadratic form

$$\begin{aligned} L^2(\hat{F}, \hat{F}_0) &= \frac{\pi T}{2} \int\int\int_{-\pi}^{\pi} \left|\hat{F}(\omega, \mathbf{u}, \mathbf{v}) - \hat{F}_0(\omega, \mathbf{u}, \mathbf{v})\right|^2 d\omega dW(\mathbf{u})\, dW(\mathbf{v}) \\ &= \sum_{j=1}^{T-1} k^2(j/p)(T-j) \int\int \left|\hat{\Gamma}_j(\mathbf{u}, \mathbf{v})\right|^2 dW(\mathbf{u})\, dW(\mathbf{v}), \end{aligned} \quad \textbf{(2.18)}$$

where the second equality follows by Parseval's identity, and $W : \mathbb{R}^d \to \mathbb{R}^+$ is a nondecreasing weighting function that weighs sets symmetric about the origin equally.[6] An example of $W(\cdot)$ is the multivariate independent $N(\mathbf{0}, \mathbf{I})$ cumulative distribution function (CDF), where $\mathbf{I}$ is a $d \times d$ identity matrix. Throughout, unspecified integrals are all taken over the support of $W(\cdot)$. We can compute the integrals over $(\mathbf{u}, \mathbf{v})$ by numerical integration. Alternatively, we can generate random draws of $\mathbf{u}$ and $\mathbf{v}$ from the prespecified distribution $W(\cdot)$, and then use the Monte Carlo simulation to approximate the integrals over $(\mathbf{u}, \mathbf{v})$. This is computationally simple and is applicable even when the dimension d is large. Note that $W(\cdot)$ need not be continuous. They can be nondecreasing step functions. This will lead to a convenient implementation of our test but it may adversely affect the power. See more discussion below.

Our test statistic for $\mathbb{H}_0$ against $\mathbb{H}_A$ is an appropriately standardized version of (2.18), namely,

$$\hat{M} = \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int\int \left|\hat{\Gamma}_j(\mathbf{u}, \mathbf{v})\right|^2 dW(\mathbf{u})\, dW(\mathbf{v}) - \hat{C}\right] \Big/ \sqrt{\hat{D}}, \quad \textbf{(2.19)}$$

where the centering factor

$$\hat{C} = \sum_{j=1}^{T-1} k^2(j/p)(T-j)^{-1} \sum_{t=|j|+1}^{T} \int\int \left|\hat{Z}_t(\mathbf{u})\right|^2 \left|\hat{\psi}_{t-j}(\mathbf{v})\right|^2 dW(\mathbf{u})\, dW(\mathbf{v}),$$

and the scaling factor

$$\hat{D} = 2\sum_{j=1}^{T-2}\sum_{l=1}^{T-2} k^2(j/p)k^2(l/p) \int\int\int\int dW(\mathbf{u}_1)\, dW(\mathbf{u}_2)\, dW(\mathbf{v}_1)\, dW(\mathbf{v}_2)$$

$$\times \left| \frac{1}{T-\max(j,l)} \sum_{t=\max(j,l)+1}^{T} \hat{Z}_t(\mathbf{u}_1)\hat{Z}_t(\mathbf{u}_2)\hat{\psi}_{t-j}(\mathbf{v}_1)\hat{\psi}_{t-l}(\mathbf{v}_2) \right|^2,$$

where $\hat{\psi}_t(\mathbf{v}) = e^{i\mathbf{v}'\mathbf{X}_t} - \hat{\varphi}(\mathbf{v})$, and $\hat{\varphi}(\mathbf{v}) = T^{-1}\sum_{t=1}^{T} e^{i\mathbf{v}'\mathbf{X}_t}$ is the empirical characteristic function of $\{\mathbf{X}_t\}$. The factors $\hat{C}$ and $\hat{D}$ are approximately the mean and variance of the quadratic form in (2.18), respectively. They have taken into account the impact of higher-order serial dependence in the generalized residual $\{Z_t(\mathbf{u})\}$. As a result, the $\hat{M}$ test is robust to conditional heteroskedasticity and time-varying higher-order conditional moments of unknown form in $\{Z_t(\mathbf{u})\}$.

In practice, $\hat{M}$ has to be calculated using numerical integration or approximated by simulation methods. This can be computationally costly when the dimension of $\mathbf{X}_t$ is large. Alternatively, one can only use a finitely large number of grid points for $\mathbf{u}$ and $\mathbf{v}$. For example, we can generate finitely many numbers of $\mathbf{u}$ and $\mathbf{v}$ from a multivariate standard normal distribution. This will dramatically reduce the computational cost but it may lead to some power loss. We will examine this issue via simulation.

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 conditional density-based approach, in the sense that our nonparametric regression estimator of CCF only involves d-dimensional smoothing, but the nonparametric joint density estimators used in the existing tests involve $2d$- and $3d$-dimensional smoothing. We expect that such dimension reduction will give better size and power performance in finite samples.

3. ASYMPTOTIC DISTRIBUTION

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

Assumption 1. (i) Assume $\{\mathbf{X}_t\}$ is a strictly stationary β-mixing process with mixing coefficient $\beta(j) = O\left(j^{-\nu}\right)$ for some constant $\nu > 12$; (ii) the marginal density $g(\mathbf{x})$ of $\mathbf{X}_t$ is bounded and Lipschitz, and the joint density $g_j(\mathbf{x}, \mathbf{y})$ of $\mathbf{X}_t$ and $\mathbf{X}_{t-j}$ is bounded.

Assumption 2. For each sufficiently large integer q, there exists a q-dependent stationary process $\{\mathbf{X}_{qt}\}$, such that $\mathrm{E}\left\|\mathbf{X}_t - \mathbf{X}_{qt}\right\|^2 \leq Cq^{-\eta}$ for some constant

$\eta \geq \frac{1}{2}$ and all large q. The random vector $\mathbf{X}_{qt}$ is measurable with respect to some sigma field, which may be different from the sigma field generated by $\{\mathbf{X}_t\}$.

Assumption 3. Let $\varphi(\mathbf{u}|\mathbf{x})$ be the CCF of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$. For each $\mathbf{u} \in \mathbb{R}^d$, $\varphi(\mathbf{u}|\mathbf{x})$is $(r+1)$th differentiable with respect to $\mathbf{x} \in \mathbb{R}^d$ and $\frac{\partial^{(r+1)}}{\partial \mathbf{x}^{(r+1)}}\varphi(\mathbf{u}|\mathbf{x})$ is Lipschitz of order α: $\left|\frac{\partial^{(r+1)}}{\partial \mathbf{x}^{(r+1)}}\varphi(\mathbf{u}|\mathbf{x}_1) - \frac{\partial^{(r+1)}}{\partial \mathbf{x}^{(r+1)}}\varphi(\mathbf{u}|\mathbf{x}_2)\right| \leq l(\mathbf{u})\|\mathbf{x}_1 - \mathbf{x}_2\|^{\alpha}$, where $0 < \alpha \leq 1$ and $\int l^2(\mathbf{u})\, dW(\mathbf{u}) < \infty$.

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

Assumption 5. (i) $k : \mathbb{R} \to [-1, 1]$ is a symmetric function that is continuous at zero and all points in $\mathbb{R}$ except for a finite number of points; (ii) $k(0) = 1$; (iii) $k(z) \leq c|z|^{-b}$ for some $b > \frac{3}{4}$ as $|z| \to \infty$.

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

Assumptions 1–3 are regularity conditions on the DGP of $\{\mathbf{X}_t\}$. Assumption 1(i) restricts the degree of temporal dependence of $\{\mathbf{X}_t\}$. We say that $\{\mathbf{X}_t\}$ is β-mixing (absolutely regular) 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 $\{\mathbf{X}_\tau : \tau = j, ..., s\}$, with $j \leq s$. Assumption 1(i) holds for many well-known processes such as stationary autoregressive moving average (ARMA) processes and a large class of processes implied by numerous nonlinear models, including bilinear, nonlinear autoregressive (AR), and autoregressive conditional heteroskedasticity (ARCH) models (Fan and Li, 1999). Ait-Sahalia, Fan, and Peng (2009), Amaro de Matos and Fernandes (2007), and Su and White (2007, 2008) also impose β-mixing conditions. Our mixing condition is weaker than those imposed in Amaro de Matos and Fernandes and in Su and White (2008). They assume a β-mixing condition with a geometric decay rate.

The proposed test is applicable to both univariate and multivariate time series with discrete or continuous distributions, or a mix of continuous and discrete data.[7] For simplicity, we just focus on the continuous case. Cases with discrete data or mixed data will be left for future research.

Assumption 2 is required only under $\mathbb{H}_0$. It assumes that a Markov process $\{\mathbf{X}_t\}$ can be approximated by a q-dependent process $\{\mathbf{X}_{qt}\}$ arbitrarily well if q is sufficiently large.[8] In fact, a Markov process can be q-dependent. Lévy (1949), Rosenblatt and Slepian (1962), Aaronson, Gilat, and Keane (1992), and Matús

(1996, 1998) provide examples of a q-dependent Markov process. Ibragimov (2007) provides the conditions that a Markov process is a q-dependent process. In this case, Assumption 2 holds trivially. Assumption 2 is not restrictive even when $\mathbf{X}_t$ is not a q-dependent process. To appreciate this, we first consider a simple AR(1) process $\{\mathbf{X}_t\}$:

$$\mathbf{X}_t = \alpha \mathbf{X}_{t-1} + \varepsilon_t, \qquad \{\varepsilon_t\} \sim \text{i.i.d.}\,(0,1)\,.$$

Define $\mathbf{X}_{qt} = \sum_{j=0}^{q} \alpha^j \varepsilon_{t-j}$, a q-dependent process. Then we have

$$\mathrm{E}\left(\mathbf{X}_t - \mathbf{X}_{qt}\right)^2 = \mathrm{E}\left(\sum_{j=q+1}^{\infty} \alpha^j \varepsilon_{t-j}\right)^2 = \frac{\alpha^{2(q+1)}}{1-\alpha}.$$

Hence Assumption 2 holds if $|\alpha| < 1$.

Another example is an ARCH(1) process $\{\mathbf{X}_t\}$:

$$\begin{cases} \mathbf{X}_t = h_t^{1/2} \varepsilon_t, \\ h_t = \alpha + \beta \mathbf{X}_{t-1}^2, \\ \varepsilon_t \sim \text{i.i.d.} N(0,1). \end{cases}$$

This is a Markov process. By recursive substitution, we have $h_t = \alpha + \alpha \sum_{j=1}^{\infty} \prod_{i=1}^{j} \beta \varepsilon_{t-i}^2$. Define $\mathbf{X}_{qt} \equiv h_{qt}^{1/2} \varepsilon_t$, where $h_{qt} \equiv \alpha + \alpha \sum_{j=1}^{q} \prod_{i=1}^{j} \beta \varepsilon_{t-i}^2$. Then $\mathbf{X}_{qt}$ is a q-dependent process, and

$$\begin{aligned} \mathrm{E}\left(\mathbf{X}_t - \mathbf{X}_{qt}\right)^2 &= \mathrm{E}\left(h_t^{1/2} - h_{qt}^{1/2}\right)^2 \le \mathrm{E}\left(h_t - h_{qt}\right) \\ &= \alpha \sum_{j=q+1}^{\infty} \prod_{i=1}^{j} \mathrm{E}\left(\beta \varepsilon_{t-i}^2\right) = \frac{\alpha \beta^{q+1}}{1-\beta}. \end{aligned}$$

Thus Assumption 2 holds if $\beta < 1$.

For the third example we consider a mean-reverting Ornstein-Uhlenbeck process $\mathbf{X}_t$:

$$d\mathbf{X}_t = \kappa\left(\theta - \mathbf{X}_t\right) dt + \sigma\, dW_t,$$

where W_t is the standard Brownian motion. This is known as Vasicek's (1977) model in the interest rate term structure literature. From the stationarity condition, we have $\mathbf{X}_t \sim N\left(\theta, \frac{\sigma^2}{2\kappa}\right)$. Define $\mathbf{X}_{qt} = \theta + \int_{t-q}^{t} \sigma e^{-\kappa(t-s)} dW_s$, which is a q-dependent process. Then Assumption 2 holds because

$$\begin{aligned} \mathrm{E}\left(\mathbf{X}_t - \mathbf{X}_{qt}\right)^2 &= \mathrm{E}\left[e^{-\kappa t}\left(\mathbf{X}_0 - \theta\right) + \int_0^{t-q} \sigma e^{-\kappa(t-s)} dW_s\right]^2 \\ &= e^{-2\kappa t}\left(\frac{\sigma^2}{2\kappa}\right) + \int_0^{t-q} \sigma^2 e^{-2\kappa(t-s)} ds \\ &= \frac{\sigma^2 e^{-2\kappa q}}{2\kappa} = o\left(q^{-\eta}\right), \quad \text{for any } \eta > 0. \end{aligned}$$

Assumption 3 provides conditions on the CCF of $\mathbf{X}_t$. As the CCF is the Fourier transform of the transition density, we can easily translate the conditions on the CCF into the conditions on the transition density $p(\mathbf{y}|\mathbf{x})$. In particular, Assumption 3 holds if for each $\mathbf{y} \in \mathbb{R}^d$, $p(\mathbf{y}|\mathbf{x})$ is $(r+1)$th differentiable with respect to $\mathbf{x} \in \mathbb{R}^d$, and $\frac{\partial^{(r+1)}}{\partial \mathbf{x}^{(r+1)}} p(\mathbf{y}|\mathbf{x})$ satisfies the Lipschitz condition of order α: $\left| \frac{\partial^{(r+1)}}{\partial \mathbf{x}^{(r+1)}} p(\mathbf{y}|\mathbf{x}_1) - \frac{\partial^{(r+1)}}{\partial \mathbf{x}^{(r+1)}} p(\mathbf{y}|\mathbf{x}_2) \right| \leq l(\mathbf{y}) \|\mathbf{x}_1 - \mathbf{x}_2\|^{\alpha}$, where $0 < \alpha \leq 1$ and $\int\int e^{2i\mathbf{u}'\mathbf{y}} l^2(\mathbf{y}) d\mathbf{y} dW(\mathbf{u}) < \infty$. Assumption 4 imposes regularity conditions on the kernel function used in local polynomial regression estimation. The same assumption has been imposed by Masry (1996a) and Ait-Sahalia et al. (2009). The condition on the boundedness and the compact support of $K(\cdot)$ is imposed for the brevity of proofs and could be removed at the cost of a more tedious proof.[9]

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

Assumption 6 imposes mild conditions on the prespecified weighting function $W(\cdot)$. Any CDF with finite fourth moments satisfies Assumption 6. Note that $W(\cdot)$ need not be continuous. This provides a convenient way to implement our tests, because we can avoid relatively high dimensional numerical integrations by using finitely many numbers of grid points for $\mathbf{u}$ and $\mathbf{v}$.

We now state the main result of this paper.

THEOREM 1. *Suppose Assumptions 1–6 hold, and* $p = cT^{\lambda}$ *for* $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$ *and* $0 < c < \infty$, $h = cT^{-\delta}$, $\delta \in \left(\frac{2-\lambda}{4(r+1)}, \min(\frac{\lambda \nu}{2d}, \frac{1-\lambda}{d}) \right)$. *Then under* $\mathbb{H}_0$, $\hat{M} \to^d N(0,1)$ *as* $T \to \infty$.

As an important feature of $\hat{M}$, the use of the nonparametrically estimated generalized residual $\hat{Z}_t(\mathbf{u})$ in place of the true unobservable residual $Z_t(\mathbf{u})$ has no impact on the limit distribution of $\hat{M}$. One can proceed as if the true CCF $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$ were known and equal to the nonparametric estimator $\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$. The reason is that by choosing suitable bandwidth h and lag order p, the convergence rate of the nonparametric CCF estimator $\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$ is faster than that of the nonparametric estimator $\hat{F}(\omega, \mathbf{u}, \mathbf{v})$ to $F(\omega, \mathbf{u}, \mathbf{v})$. Consequently, the limiting distribution of $\hat{M}$ is solely determined by $\hat{F}(\omega, \mathbf{u}, \mathbf{v})$, and replacing $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$

by $\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$ has no impact on the asymptotic distribution of $\hat{M}$ under $\mathbb{H}_0$. The impact of the first-stage estimation comes from two sources—bias and variance—and we have to balance them. The dimension d affects the variance but not the bias. For given T and h, the variance increases with the dimension and, consequently, a smaller dimension allows for a bigger feasible range of δ. The dimension d has no direct impact on λ, as the frequency domain estimation is used for the one-dimensional generalized spectrum $F(\omega, \mathbf{u}, \mathbf{v})$, no matter how big the dimension of $\mathbf{X}_t$ is. However, since we need to balance the convergence speeds of h and p, the dimension d has an indirect impact on p. The smaller the dimension is, the bigger the feasible range of λ would be.

Although the use of $\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$ has no impact on the limit distribution of the $\hat{M}$ test, it may have substantial impact on its finite sample size performance. To overcome such adverse impact, we will use Horowitz's (2003) nonparametric smoothed transition density-based bootstrap procedure to obtain the critical values of the test in finite samples. See more discussion in Section 5 below.

4. ASYMPTOTIC POWER

Our test is derived without assuming a specific alternative to $\mathbb{H}_0$. To get insights into the nature of the alternatives that our test is able to detect, we now examine the asymptotic power behavior of $\hat{M}$ under $\mathbb{H}_A$ in (2.2).

THEOREM 2. *Suppose Assumptions 1 and 3–6 hold, and* $p = cT^{\lambda}$ *for* $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$ *and* $0 < c < \infty$, $h = cT^{-\delta}$, $\delta \in \left(\frac{2-\lambda}{4(r+1)}, \min(\frac{\lambda\nu}{2d}, \frac{1}{d})\right)$. *Then under* $\mathbb{H}_A$, *and as* $T \to \infty$,

$$\frac{p^{1/2}}{T}\hat{M} \to^{P} \frac{1}{\sqrt{D}} \sum_{j=1}^{\infty} \int\int \left|\Gamma_j(\mathbf{u}, \mathbf{v})\right|^2 dW(\mathbf{u})\, dW(\mathbf{v})$$

$$= \frac{\pi}{2\sqrt{D}} \int\int\int_{-\pi}^{\pi} |F(\omega, \mathbf{u}, \mathbf{v}) - F_0(\omega, \mathbf{u}, \mathbf{v})|^2 d\omega dW(\mathbf{u})\, dW(\mathbf{v}),$$

where

$$D = 2\int_0^{\infty} k^4(z)\, dz \int\int |\Sigma_0(\mathbf{u}_1, \mathbf{u}_2)|^2 dW(\mathbf{u}_1)$$

$$dW(\mathbf{u}_2) \sum_{j=-\infty}^{\infty} \int\int \left|\Omega_j(\mathbf{v}_1, \mathbf{v}_2)\right|^2 dW(\mathbf{v}_1)\, dW(\mathbf{v}_2),$$

$$\Omega_j(\mathbf{u}, \mathbf{v}) = \operatorname{cov}\left(e^{i\mathbf{u}'\mathbf{X}_t}, e^{i\mathbf{v}'\mathbf{X}_{t-|j|}}\right) \text{ and } \Sigma_0(\mathbf{u}, \mathbf{v}) = \operatorname{cov}[Z_t(\mathbf{u}), Z_t(\mathbf{v})].$$

The restriction on h in Theorem 2 is weaker than that in Theorem 1, as we allow for a slower convergence rate of the first-stage nonparametric estimation. The function $L(\omega, \mathbf{u}, \mathbf{v})$ is the generalized spectral density of the process $\{\mathbf{X}_t\}$, which is first introduced in Hong (1999) in a univariate context. It captures temporal dependence in $\{\mathbf{X}_t\}$. The dependence of the constant D on $L(\omega, \mathbf{u}, \mathbf{v})$ is due to the fact that the conditioning variable $\{e^{i\mathbf{v}'\mathbf{X}_{t-|j|}}\}$ is a time series process. This suggests that if the time series $\{\mathbf{X}_t\}$ is highly persistent, it may be more difficult to detect violation of the Markov property because the constant D will be larger.

Following reasoning analogous to Bierens (1982) and Stinchcombe and White (1998), we have that for $j > 0$, $\Gamma_j(\mathbf{u},\mathbf{v}) = 0$ for all $\mathbf{u},\mathbf{v} \in \mathbb{R}^d$ if and only if $\mathrm{E}\left[Z_t(\mathbf{u})|\mathbf{X}_{t-j}\right] = 0$ a.s. for all $\mathbf{u} \in \mathbb{R}^d$. Thus, the generalized covariance function $\Gamma_j(u,v)$ can capture various departures from the Markov property in every conditional moment of $\mathbf{X}_t$ in view of the Taylor series expansion in (2.7). Suppose $\mathrm{E}\left[Z_t(\mathbf{u})|\mathbf{X}_{t-j}\right] \neq 0$ at some lag $j > 0$. Then we have $\int\int\left|\Gamma_j(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\, dW(\mathbf{v}) > 0$ for any weighting function $W(\cdot)$ that is positive, monotonically increasing, and continuous, with unbounded support on $\mathbb{R}^d$. Consequently, $P[\hat{M} > C(T)] \to 1$ for any sequence of constants $\{C(T) = o(T/p^{1/2})\}$. Thus $\hat{M}$ has asymptotic unit power at any given significance level, whenever $\mathrm{E}\left[Z_t(\mathbf{u})|\mathbf{X}_{t-j}\right] \neq 0$ at some lag $j > 0$.

Thus, to ensure the consistency property of $\hat{M}$, it is important to integrate $\mathbf{u}$ and $\mathbf{v}$ over the entire domain of $\mathbb{R}^d$. When numerical integration is difficult, as is the case where the dimension d is large, one can use Monte Carlo simulation to approximate the integrals over $\mathbf{u}$ and $\mathbf{v}$. This can be obtained by using a large number of random draws from the distribution $W(\cdot)$ and then computing the sample average as an approximation to the related integral. Such an approximation will be arbitrarily accurate provided the number of random draws is sufficiently large. Alternatively, we can use a nondecreasing step function $W(\cdot)$. This avoids numerical integration or Monte Carlo simulation, but the power of the test may be affected. In theory, the consistency property will not be preserved if only a finite number of grid points of $\mathbf{u}$ and $\mathbf{v}$ are used, and the power of the test may depend on the choice of grid points for $\mathbf{u}$ and $\mathbf{v}$.

On the other hand, Theorem 2 implies that the $\hat{M}$ test can check departure from the Markov property at any lag order $j > 0$, as long as the sample size T is sufficiently large. This is achieved because $\hat{M}$ includes an increasing number of lags as the sample size $T \to \infty$. Usually the use of a large number of lags would lead to the loss of a large number of degrees of freedom. Fortunately this is not the case with the $\hat{M}$ test, thanks to the downward weighting of $k^2(\cdot)$ for higher-order lags.

As revealed by the Taylor series expansion in (2.7), our test, which is based on the MDS characterization in (2.6), essentially checks departures from the Markov property in every conditional moment. When $\hat{M}$ rejects the Markov property, one may be further interested in what causes the rejection. To gauge possible sources of the violation of the Markov property, we can construct a sequence of tests based on the derivatives of the nonparametric regression residual $Z_t(\mathbf{u})$ at the origin $\mathbf{0}$:

$$\frac{\partial^{|\mathbf{m}|}}{\partial u_1^{m_1}\cdots\partial u_d^{m_d}}\mathrm{E}\left[Z_t(\mathbf{u})|\mathcal{I}_{t-1}\right]_{\mathbf{u}=\mathbf{0}} = \mathrm{E}(X_{1t}^{m_1}\cdots X_{dt}^{m_d}|\mathcal{I}_{t-1}) - \mathrm{E}(X_{1t}^{m_1}\cdots X_{dt}^{m_d}|\mathbf{X}_{t-1}) = 0,$$

where the order of derivatives $|\mathbf{m}| = \Sigma_{a=1}^{d} m_a$, and $\mathbf{m} = (m_1, ..., m_d)'$, and $m_a \geq 0$ for all $a = 1, ..., d$. For the univariate time series (i.e., $d = 1$), the choices of $\mathbf{m} = 1, 2, 3, 4$ correspond to tests for departures of the Markov property in the first four conditional moments, respectively. For each $\mathbf{m}$, the resulting test statistic is given by

$$\hat{M}(\mathbf{m}) = \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int \left| \hat{\Gamma}_j^{(\mathbf{m},\mathbf{0})}(\mathbf{0}, \mathbf{v}) \right|^2 dW(\mathbf{v}) - \hat{C}(\mathbf{m}) \right] \Big/ \sqrt{\hat{D}(\mathbf{m})}, \tag{4.1}$$

where $\hat{\Gamma}_j^{(\mathbf{m},\mathbf{0})}(\mathbf{0}, \mathbf{v})$ is the sample analogue of the derivative of the generalized cross-covariance function

$$\Gamma_j^{(\mathbf{m},\mathbf{0})}(\mathbf{0}, \mathbf{v}) = \operatorname{cov}\left\{ \prod_{a=1}^{d} (iX_{at})^{m_a} - \mathrm{E}\left[\prod_{a=1}^{d} (iX_{at})^{m_a} \middle| \mathbf{X}_{t-1} \right], e^{\left(i\mathbf{v}'\mathbf{X}_{t-|j|}\right)} \right\},$$

the centering and scaling factors

$$\hat{C}(\mathbf{m}) = \sum_{j=1}^{T-1} k^2(j/p) \frac{1}{T-j} \sum_{t=|j|+1}^{T} \int \left| \hat{Z}_t^{(\mathbf{m})}(\mathbf{0}) \right|^2 \left| \hat{\psi}_{t-j}(\mathbf{v}) \right|^2 dW(\mathbf{v}),$$

$$\hat{D}(\mathbf{m}) = 2 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/p) k^2(l/p) \int\int \times \left| \frac{1}{T - \max(j,l)} \sum_{t=\max(j,l)+1}^{T} \left| \hat{Z}_t^{(\mathbf{m})}(\mathbf{0}) \right|^2 \hat{\psi}_{t-j}(\mathbf{v}_1) \hat{\psi}_{t-j}(\mathbf{v}_2) \right|^2 dW(\mathbf{v}_1) dW(\mathbf{v}_2),$$

and

$$\hat{Z}_t^{(\mathbf{m})}(\mathbf{0}) = \prod_{a=1}^{d} (iX_{at})^{m_a} - \mathrm{E}\left[\prod_{a=1}^{d} (iX_{at})^{m_a} | \mathbf{X}_{t-1} \right].$$

These derivative tests may provide additional useful information on the possible sources of the violation of the Markov property. On the other hand, some economic theories only have implications for the Markov property in certain moments, and our derivative tests are suitable to test these implications. For example, Hall (1978) shows that a rational expectation model of consumption can be characterized by the Euler equation that $\mathrm{E}\left[u'(C_{t+1}) | \mathcal{I}_t\right] = u'(C_t)$, where $u'(C_t)$ is the marginal utility of consumption C_t. This can be viewed as the Markov property in mean for the marginal utility process of consumption. The derivative test $\hat{M}(1)$ can be used to test this implication.

5. NUMERICAL RESULTS

5.1. Monte Carlo Simulations

Theorem 1 provides the null asymptotic $N(0,1)$ distribution of $\hat{M}$. Thus, one can implement our test for $\mathbb{H}_0$ by comparing $\hat{M}$ with a $N(0,1)$ critical value. However, like many other nonparametric tests in the literature, the size of $\hat{M}$ in finite samples may differ significantly from the prespecified asymptotic significance level. Our analysis suggests that the asymptotic theory may not work well even for relatively large sample sizes, because the asymptotically negligible higher-order terms in $\hat{M}$ are close in order of magnitude to the dominant U-statistic that determines the limit distribution of $\hat{M}$. In particular, the first-stage smoothed nonparametric regression estimation for $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$ may have substantial adverse effect on the size of $\hat{M}$ in finite samples. Indeed, our simulation study shows that $\hat{M}$ displays severe underrejection under $\mathbb{H}_0$. We examine the finite sample performance of an infeasible $\hat{M}$ test by replacing the estimated generalized residual $\hat{Z}_t(\mathbf{u})$ with the true generalized residual $Z_t(\mathbf{u})$. We find that the size of the infeasible test is reasonable. This experiment suggests that the underrejection of $\hat{M}$ is mainly due to the impact of the first-stage nonparametric estimation of CCF, which has a rather slow convergence rate. Similar problems are also documented by Skaug and Tjøstheim (1993, 1996), Hong and White, (2005), and Fan, Li, and Min (2006) in other contexts.

To overcome this problem, we use Horowitz's (2003) smoothed nonparametric conditional density bootstrap procedure to approximate the null finite-sample null distribution of $\hat{M}$ more accurately. The basic idea is to use a smoothed nonparametric transition density estimator (under $\mathbb{H}_0$) to generate bootstrap samples. Specifically, it involves the following steps:

Step (i). To obtain a bootstrap sample $\mathcal{X}^b \equiv \{\mathbf{X}_t^b\}_{t=1}^T$, draw $\mathbf{X}_1^b$ from the smoothed unconditional kernel density

$$\hat{g}(\mathbf{x}) = \frac{1}{T}\sum_{s=2}^{T}\mathbf{K}_h(\mathbf{x}-\mathbf{X}_{s-1})$$

and $\{\mathbf{X}_t^b\}_{t=2}^T$ recursively from the smoothed conditional kernel density

$$\hat{g}(\mathbf{x}|\mathbf{X}_{t-1}^b) = \frac{\frac{1}{T}\sum_{s=2}^{T}\mathbf{K}_h(\mathbf{x}-\mathbf{X}_s)\mathbf{K}_h\left(\mathbf{X}_{t-1}^b-\mathbf{X}_{s-1}\right)}{\frac{1}{T}\sum_{s=2}^{T}\mathbf{K}_h\left(\mathbf{X}_{t-1}^b-\mathbf{X}_{s-1}\right)}, \tag{5.1}$$

where $\mathbf{K}(\cdot)$ and h are the same as those used in $\hat{M}$.

Step (ii). Compute a bootstrap statistic $\hat{M}^b$ in the same way as $\hat{M}$, with $\mathcal{X}^b$ replacing $\mathcal{X}=\{\mathbf{X}_t\}_{t=1}^T$. The same $\mathbf{K}(\cdot)$ and h are used in $\hat{M}$ and $\hat{M}^b$.

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

Step (iv). Compute the bootstrap p-value $p_b \equiv B^{-1}\sum_{l=1}^{B} \mathbf{1}(\hat{M}_l^b > \hat{M})$ for a sufficiently large B.

We suggest using the same kernel $\mathbf{K}(\cdot)$ and the same bandwidth h in computing $\hat{g}(\mathbf{x}|\mathbf{X}_{t-1})$, $\hat{M}$ and $\hat{M}^b$. This is not necessary, but it delivers a simpler test procedure.[10] Smoothed nonparametric bootstraps have been used to improve finite sample performance in hypothesis testing. For example, Su and White (2007, 2008) apply Paparoditis and Politis's (2000) procedure in testing for conditional independence, and Amaro de Matos and Fernandes (2007) use Horowitz's (2003) Markov conditional bootstrap procedure in testing for the Markov property. Paparoditis and Politis's (2000) procedure is similar to Horowitz's, except that Paparoditis and Politis (2000) generate bootstrap samples from $\hat{g}(\mathbf{x}|\mathbf{X}_{t-1})$ and Horowitz generates bootstrap samples from $\hat{g}(\mathbf{x}|\mathbf{X}_{t-1}^b)$. Both methods can be applied to our test, although Horowitz's procedure is more computationally expensive.[11] When Paparoditis and Politis's (2000) method is used, the bootstrap sample $\{\mathbf{X}_t^b\}_{t=1}^T$ is an i.i.d. sequence conditional on the original sample $\mathcal{X}$ and hence it is Markov conditional on $\mathcal{X}$. Following an analogous proof of Theorem 4.1 of Su and White (2008), we can show that conditional on $\mathcal{X}$, $\hat{M}^b \to^d N(0,1)$ as $T \to \infty$. The proof is similar to but simpler than that of Theorem 1 in Section 3 due to the fact that $\{\mathbf{X}_t^b\}_{t=1}^T$ is i.i.d. conditional on $\mathcal{X}$. More specifically, we can first show that the estimation uncertainty in the first-stage nonparametric estimation has no impact asymptotically. Then, by applying Brown's (1971) central limit theorem, we can derive the asymptotic normality of $\hat{M}^b$ conditional on $\mathcal{X}$. On the other hand, the proof of consistency with the Horowitz approach is much more involved. We conjecture that following an analogous proof of Theorem 3.4 of Paparoditis and Politis (2002), we can show that conditional on $\mathcal{X}$, $\mathbf{X}_t^b$ is a so-called ρ-mixing process with a geometric decay rate. Then by applying a suitable central limit theorem of the degenerate U-statistics (e.g., Gao and Hong, 2008, Thm 2.1), the asymptotic normality of $\hat{M}^b$ conditional on $\{\mathbf{X}_t\}_{t=1}^T$ may be obtained.

The consistency of the smoothed bootstrap does not indicate the degree of improvement of the smoothed bootstrap upon the asymptotic distribution. Since $\hat{M}$ is asymptotically pivotal, it is possible that $\hat{M}^b$ can achieve reasonable accuracy in finite samples. We shall examine the performance of the smoothed bootstrap in our simulation study.

We shall compare the finite sample performance of our $\hat{M}$ test with Su and White's (SW) (2007, 2008) CCF-based test and Hellinger metric test for conditional independence.[12] To examine the size of the tests under $\mathbb{H}_0$, we consider two Markov DGPs:

DGP S1 [AR(1)]: $\quad X_t = 0.5X_{t-1} + \varepsilon_t,$

DGP S2 [ARCH(1)]: $\begin{cases} X_t = h_t^{1/2}\varepsilon_t \\ h_t = 0.1 + 0.1X_{t-1}^2, \end{cases}$

where $\varepsilon_t \sim$ i.i.d.$N(0,1)$.

To examine the power of the tests using the smoothed bootstrap, we consider the following non-Markovian DGPs:

DGP P1 [MA(1)]: $X_t = \varepsilon_t + 0.5\varepsilon_{t-1},$

DGP P2 [GARCH(1,1)]: $\begin{cases} X_t = h_t^{1/2}\varepsilon_t \\ h_t = 0.1 + 0.2X_{t-1}^2 + 0.7h_{t-1}, \end{cases}$

DGP P3 [GARCH-in-Mean]: $\begin{cases} X_t = 0.3 + 0.5h_t + z_t \\ z_t = h_t^{1/2}\varepsilon_t \\ h_t = 0.1 + 0.2X_{t-1}^2 + 0.7h_{t-1}, \end{cases}$

DGP P4 [Markov Chain Regime-Switching]: $X_t = \begin{cases} 0.7X_{t-1} + \varepsilon_t, & \text{if } S_t = 0, \\ -0.3X_{t-1} + \varepsilon_t, & \text{if } S_t = 1, \end{cases}$

DGP P5 [Markov Chain Regime-Switching ARCH]: $\begin{cases} X_t = \begin{cases} \sqrt{h_t}\varepsilon_t, & \text{if } S_t = 0, \\ 3\sqrt{h_t}\varepsilon_t, & \text{if } S_t = 1, \end{cases} \\ h_t = 0.1 + 0.3X_{t-1}^2, \end{cases}$

where $\varepsilon_t \sim$ i.i.d.$N(0,1)$, and in DGPs P4 and P5, S_t is a latent state variable that follows a two-state Markov chain with transition probabilities $P(S_t = 1|S_{t-1} = 0) = P(S_t = 0|S_{t-1} = 1) = 0.9$. DGPs P4 and P5 are the Markov Chain Regime-Switching model and Markov Chain Regime-Switching ARCH model proposed by Hamilton (1989) and Hamilton and Susmel (1994), respectively. They can capture the state-dependent behaviors in time series. The introduction of S_t changes the Markov property of AR(1) and ARCH(1) processes. The knowledge of X_{t-1} is not sufficient to summarize all relevant information in $\mathcal{I}_{t-1}$ that is useful to predict the future behavior of X_t. The departure from the Markov property comes from the conditional mean in DGPs P1 and P4, from the conditional variance in DGPs P2 and P5, and from both the conditional mean and conditional variance in DGP P3.

Throughout, we consider three sample sizes: $T = 100, 250, 500$. For each DGP we first generate $T + 100$ observations and then discard the first 100 to mitigate the impact of the initial values. To examine the bootstrap sizes and powers of the tests, we generate 500 realizations of the random sample $\{X_t\}_{t=1}^T$, using the Gauss Windows version random number generator. We use $B = 100$ bootstrap iterations for each simulation iteration. To reduce computational costs of our $\hat{M}$ test, we generate $\mathbf{u}$ and $\mathbf{v}$ from an $N(0,1)$ distribution, with each $\mathbf{u}$ and $\mathbf{v}$ having 30 symmetric grid points in $\mathbb{R}$, respectively.[13] We use the Bartlett kernel in (2.15), which has bounded support and is computationally efficient. Our simulation experience suggests that the choices of $W(\cdot)$ and $k(\cdot)$ have little impact on both the size and power of the tests.[14] Like Hong (1999), we use a data-driven $\hat{p}$ via a plug-in method that minimizes the asymptotic integrated mean squared error of the generalized spectral density estimator $\hat{F}(\omega, \mathbf{x}, \mathbf{y})$, with the Bartlett kernel $k(\cdot)$ used in some preliminary generalized spectral estimators. To examine the sensitivity of the choice of a preliminary bandwidth $\bar{p}$ on the size and power of the $\hat{M}$ test, we consider $\bar{p}$ in the range of 5 to 20. We use the Gaussian kernel for $\mathbf{K}(\cdot)$. For simplicity, we choose $h = \hat{S}_X T^{-1/4.5}$, where $\hat{S}_X$ is the sample standard deviation of $\{X_t\}_{t=1}^{T}$.[15] We compare the proposed test with Su and White's (2007, 2008) tests, applied to the present context to check whether X_t is independent of X_{t-2} conditional on X_{t-1}. Following Su and White (2008, 2007), we choose a fourth-order kernel $\mathbf{K}(u) = (3-u^2)\varphi(u)/2$, where $\varphi(\cdot)$ is the $N(0,1)$ density function, $h = T^{-1/8.5}$ for the nonparametric estimation of their Hellinger metric test SW_a, $h_1 = h_1^* T^{1/10} T^{-1/6}$ and $h_2 = h_2^* T^{1/9} T^{-1/5}$ for their CCF-based test SW_b, where h_1^* and h_2^* are the least-squares cross-validated bandwidths for estimating the conditional expectations of X_t given (X_{t-1}, X_{t-2}) and X_{t-1}, respectively, and $b = T^{-1/5}$ for the bootstrap.

Table 1 reports the bootstrap sizes and powers of $\hat{M}$, SW_a, and SW_b at the 10% and 5% levels under DGPs S1–S2 and P1–P5. The $\hat{M}$ test has reasonable sizes under the DGPs S1 and S2 at both 10% and 5% levels. Under DGP S1 (AR(1)) the empirical levels of $\hat{M}$ are very close to the nominal levels, especially at the 5% level. When $T = 100$, $\hat{M}$ tends to overreject a little under DGP S2 (ARCH(1)), but the overrejection is not excessive, and it improves as T increases. The sizes of $\hat{M}$ are not very sensitive to the choice of the preliminary lag order $\bar{p}$. The smoothed bootstrap procedure has reasonable sizes in small samples. We note that the rejection rate of SW_a decreases monotonically under DGP S1 and reaches 2.8% at the 5% level when $T = 500$, but SW_b has good sizes under both DGPs.

Under DGPs P1–P5, X_t is not Markov, and our test has reasonable power. Under DGPs P1 and P4 (MA(1) and Markov Chain Regime-Switching), our test dominates SW_a and SW_b for all sample sizes considered. Interestingly, SW_a and SW_b have nonmonotonic power against DGP P4, and their rejection rates only reach 10.4% and 7.2%, respectively, at the 5% level when $T = 500$. In contrast, the power of $\hat{M}$ is around 50% at the 5% level when $T = 500$. Under DGPs P2, P3, and P5 (GARCH(1,1), Markov Chain Regime-Switching ARCH, and GARCH-in-Mean), SW_a and SW_b perform slightly better in small samples, but the power of our $\hat{M}$ test increases more quickly with T, and our test outperforms SW_a and SW_b when $T = 500$, which demonstrates the nice feature of our frequency domain approach. The relative ranking between SW_a and SW_b does not display a very clear pattern, but SW_b is more powerful under DGPs P1–P3.

In summary, the new $\hat{M}$ test with the smoothed bootstrap procedure delivers reasonable size and omnibus power against various non-Markov alternatives in small samples. It performs well relative to two existing tests SW_a and SW_b in many cases.

5.2. Application to Financial Data

As documented by Hong and Li (2005), such popular spot interest rate continuous-time models as Vasicek (1977), Cox et al. (1985), Chan, Karolyi, Longstaff, and Sanders (1992), Ait-Sahalia (1996), and Ahn and Gao (1999) are all strongly rejected with real interest rate data. They cannot capture the full dynamics of

TABLE 1. Size and power of the test

	$T=100$					$T=250$					$T=500$				
	$\hat{M}$			SW_a	SW_b	$\hat{M}$			SW_a	SW_b	$\hat{M}$			SW_a	SW_b
lag	10	15	20			10	15	20			10	15	20		
Size															
DGP S1: AR(1)															
10%	.066	.088	.090	.080	.112	.094	.098	.096	.072	.090	.088	.086	.092	.058	.088
5%	.042	.042	.048	.040	.064	.036	.044	.048	.036	.050	.044	.048	.044	.028	.048
DGP S2: ARCH(1)															
10%	.116	.122	.126	.164	.102	.082	.094	.098	.138	.100	.094	.092	.092	.086	.100
5%	.070	.064	.066	.102	.040	.046	.040	.040	.078	.058	.048	.048	.050	.050	.050
Power															
DGP P1: MA(1)															
10%	.278	.262	.236	.128	.138	.444	.424	.390	.156	.260	.718	.674	.616	.252	.360
5%	.156	.144	.136	.076	.072	.328	.300	.256	.098	.166	.622	.552	.508	.158	.264
DGP P2: GARCH(1,1)															
10%	.172	.158	.150	.218	.234	.224	.242	.258	.210	.284	.440	.452	.446	.310	.372
5%	.084	.086	.078	.150	.160	.154	.166	.162	.136	.206	.274	.300	.296	.216	.234
DGP P3: GARCH-in-Mean															
10%	.164	.168	.174	.188	.206	.348	.360	.366	.246	.340	.628	.648	.668	.362	.508
5%	.090	.102	.088	.114	.120	.224	.234	.246	.162	.246	.490	.540	.536	.254	.362
DGP P4: Markov Regime-Switching															
10%	.244	.214	.202	.190	.134	.442	.384	.348	.180	.150	.666	.612	.578	.164	.120
5%	.156	.148	.140	.114	.078	.302	.270	.252	.094	.070	.550	.494	.458	.104	.072
DGP P5: Markov Chain Regime-Switching ARCH															
10%	.174	.152	.154	.100	.188	.328	.320	.298	.364	.288	.626	.594	.590	.560	.388
5%	.098	.086	.082	.042	.112	.204	.202	.198	.262	.162	.496	.478	.456	.448	.240

Notes: (i) $\hat{M}$ is our proposed omnibus test, given in (2.19); SW_a and SW_b are Su and White's (2008) Hellinger metric test and Su and White's (2007) characteristic function-based test, respectively; (ii) 500 iterations and 100 bootstrap iterations for each simulation iteration.

the spot interest rates. Although works are still going on to add the richness of model specification in terms of jumps and functional forms, the models proposed continue to be a Markov process. In fact, the firm rejection of a continuous-time model could be due to the violation of the Markov property, as speculated by Hong and Li. If this is indeed the case, one should not attempt to look for flexible functional forms within the class of Markov models. On the other hand, as discussed earlier, an important conclusion of the asymmetric information microstructure models (e.g., Easley and O'Hara, 1987, 1992) is that asset price sequences do not follow a Markov process. It is interesting to check whether real stock prices are consistent with this conjecture.

We apply our test to three important financial time series—stock prices, interest rates, and foreign exchange rates—and compare it with SW_a and SW_b. We use the Standard & Poor's 500 price index, 7-day Eurodollar rate, and Japanese yen, obtained from Datastream. The data are weekly series from 1 January 1988 to 31 December 2006. The weekly series are generated by selecting Wednesdays series (if a Wednesday is a holiday then the preceding Tuesday is used), which all have 991 observations. The use of weekly data avoids the so-called weekend effect, as well as other biases associated with nontrading, asynchronous rates, and so on, which are often present in higher-frequency data. To examine the sensitivity of our conclusion to the possible structural changes, we consider two subsamples: 1 January 1988 to 31 December 1997, for a total of 521 observations, and 1 January 1998 to 31 December 2006, for a total of 470 observations. Figure 1 provides the time series plots and Table 2 reports some descriptive statistics. The augmented Dickey-Fuller test indicates that there exists a unit root in all three level series but not in their first differenced series. Therefore, as is a standard practice, we use S&P 500 log returns, 7-day Eurodollar rate changes, and Japanese yen log returns. To check possible structural changes, we use Inoue's (2001) Kolmogorov-Smirnov (KS) test for the stability of stationary distributions.[16] Table 2 shows that we are unable to reject the distribution stability hypothesis for all series in both sample periods at the 5% level, and we are only able to reject the distribution stability hypothesis for 7-day Eurodollar rate changes for the full sample at the 10% level. On the other hand, to check the robustness of our test to possible structural breaks, we apply our test to an AR(1) model with structural break in a simulation study (results are availabe upon requests). This DGP is Markov, but there exists a structural break. Our test does not overreject the null Markov hypothesis. This suggests that our test may be robust to some forms of structural breaks in practice.

Table 3 reports the test statistics and bootstrap p-values of our test, SW_a, and SW_b. The bootstrap p-values, based on $B = 500$ bootstrap iterations, are computed as described in Section 5.1. For all sample periods considered, the bootstrap p-values of our test statistics are quite robust to the choice of the preliminary lag order $\bar{p}$. For the whole sample and the subsample of 1998 to 2006, we find strong evidence against the Markov property for S&P 500 returns, 7-day Eurodollar rate changes, and Japanese yen returns: All bootstrap p-values of our test are smaller

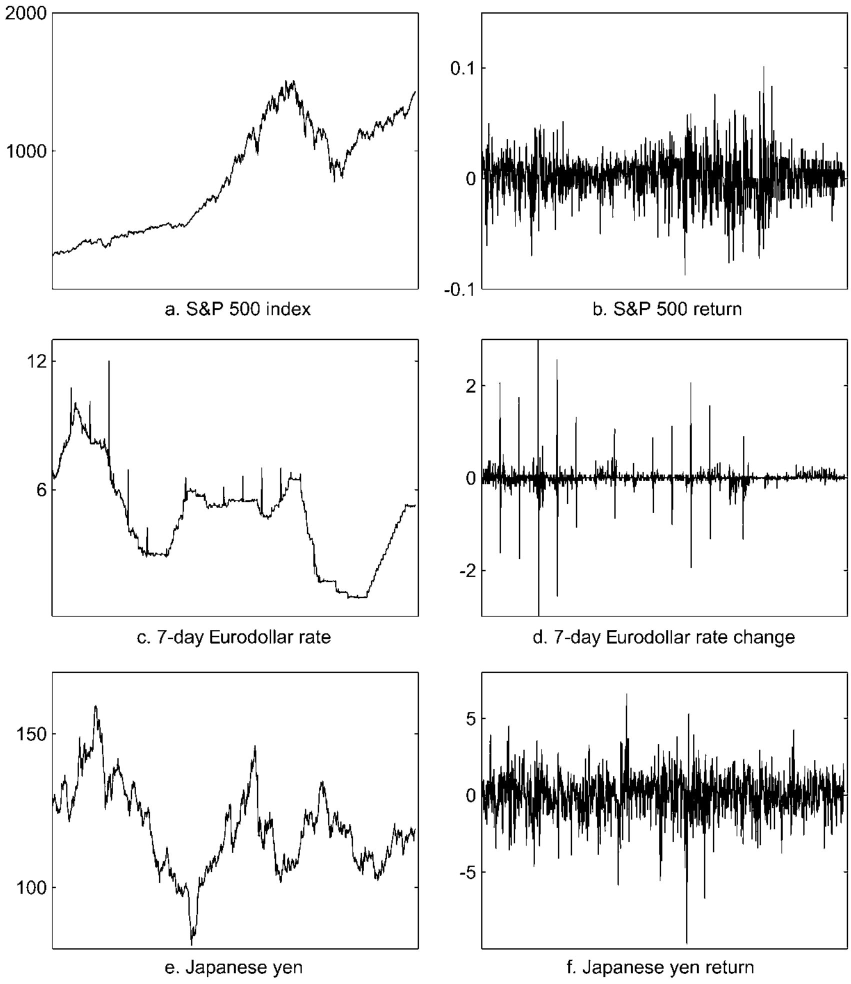

FIGURE 1. Financial time series plots.

than 5%. For the subsample of 1988 to 1997, we only reject the Markov property of 7-day Eurodollar rate changes at the 5% level. The results of SW_a and SW_b are mixed, and there seems no clear pattern of these two tests. For example, at the 10% level, SW_a is only able to reject the Markov property of S&P 500 returns and 7-day Eurodollar rate changes from 1998 to 2006, and SW_b is only able to reject that of S&P 500 returns from 1988 to 2006 and 7-day Eurodollar rate changes from 1988 to 2006 and 1988 to 1997.

To gauge possible sources of the violation of the Markov property, we also implement derivative tests $\hat{M}(m)$, $m = 1, 2, 3, 4$, as described in Section 4. Tests and their results are reported in Table 4. We first consider S&P 500 returns. A

TABLE 2. Descriptive statistics for S&P 500, interest rate, and exchange rate

	01/01/1988 – 31/12/2006			01/01/1988 – 31/12/1997			01/01/1998 – 31/12/2006		
	S&P	Eurodollar	JY	S&P	Eurodollar	JY	S&P	Eurodollar	JY
Sample size	991	991	991	521	521	521	470	470	470
Mean	0.0017	−0.0017	−0.0001	0.0025	−0.0025	0.0000	0.0008	−0.0004	−0.0002
Std	0.0209	0.3272	0.0145	0.0179	0.4087	0.0146	0.0238	0.2019	0.0145
ADF	−0.58	−1.19	−2.07	2.10	−1.00	−1.19	−1.78	−0.72	−2.42
	(0.8728)	(0.6808)	(0.2550)	(0.9999)	(0.7532)	(0.6786)	(0.3896)	(0.8395)	(0.1362)
KS	0.2828	0.0707	0.3939	0.6364	0.1010	0.1818	0.1818	0.1414	0.2828

Notes: ADF denotes the augmented Dickey-Fuller test; KS denotes the Kolmogorov-Smirnov test for the stability of stationary distributions proposed by Inoue (2001).

TABLE 3. Markov test for S&P 500, interest rate, and exchange rate

	S&P 500		7-day Eurodollar rate		Japanese yen	
lag	Statistics	p-values	Statistics	p-values	Statistics	p-values
$\hat{M}$			01/01/1988 – 31/12/2006			
10	0.86	0.0160	0.75	0.0080	1.34	0.0000
11	0.86	0.0160	0.98	0.0040	1.39	0.0000
12	0.89	0.0160	1.16	0.0040	1.52	0.0000
13	0.95	0.0160	1.35	0.0040	1.65	0.0000
14	1.01	0.0160	1.58	0.0020	1.76	0.0000
15	1.05	0.0180	1.79	0.0020	1.85	0.0000
16	1.07	0.0180	1.99	0.0020	1.94	0.0000
17	1.09	0.0200	2.22	0.0020	2.01	0.0000
18	1.11	0.0180	2.48	0.0020	2.08	0.0000
19	1.12	0.0180	2.74	0.0020	2.15	0.0000
20	1.13	0.0180	2.97	0.0000	2.21	0.0000
SW_a	0.79	0.1680	−4.61	0.9920	0.09	0.4600
SW_b	0.36	0.0940	0.21	0.0520	−0.85	0.5700
$\hat{M}$			01/01/1988 – 31/12/1997			
10	−1.39	0.5940	0.25	0.0100	−0.35	0.0980
11	−1.39	0.6120	0.30	0.0100	−0.35	0.1060
12	−1.35	0.6080	0.34	0.0080	−0.27	0.1020
13	−1.30	0.5900	0.38	0.0060	−0.21	0.0980
14	−1.25	0.5840	0.41	0.0080	−0.16	0.0980
15	−1.20	0.5780	0.45	0.0080	−0.12	0.1040
16	−1.15	0.5600	0.49	0.0080	−0.08	0.1040
17	−1.08	0.5260	0.51	0.0080	−0.04	0.1060
18	−1.02	0.5080	0.54	0.0100	−0.01	0.1100
19	−0.96	0.4860	0.57	0.0100	0.03	0.1100
20	−0.91	0.4640	0.62	0.0080	0.06	0.1100
SW_a	−0.36	0.6540	−4.85	0.9900	0.16	0.3640
SW_b	−0.14	0.1680	0.07	0.0700	0.03	0.1100
$\hat{M}$			01/01/1998 – 31/12/2006			
10	1.68	0.0080	0.34	0.0100	0.71	0.0100
11	1.88	0.0060	0.74	0.0040	0.76	0.0120
12	2.06	0.0040	1.08	0.0000	0.82	0.0140
13	2.22	0.0020	1.36	0.0000	0.88	0.0140
14	2.36	0.0020	1.62	0.0000	0.94	0.0140
15	2.48	0.0020	1.87	0.0000	0.98	0.0140
16	2.58	0.0020	2.09	0.0000	1.02	0.0100
17	2.66	0.0000	2.27	0.0000	1.06	0.0100
18	2.74	0.0000	2.44	0.0000	1.09	0.0100
19	2.81	0.0000	2.60	0.0000	1.11	0.0120
20	2.88	0.0000	2.75	0.0000	1.14	0.0120
SW_a	1.12	0.0960	1.50	0.0180	0.63	0.2160
SW_b	−0.07	0.1520	−0.18	0.1740	−1.28	0.8600

Notes: (i) $\hat{M}$ is our proposed omnibus test,given in (2.19); SW_a and SW_b are Su and White's (2008) Hellinger metric test and Su and White's (2007) characteristic function based test respectively; (ii) 500 bootstrap iterations.

TABLE 4. Derivative tests for S&P 500, interest rate, and exchange rate

	$\hat{M}(1)$	$\hat{M}(2)$	$\hat{M}(3)$	$\hat{M}(4)$
S&P 500 1988–2006	0.4220	0.2380	0.2160	0.2220
S&P 500 1988–1997	0.6320	0.0140	0.2680	0.0280
S&P 500 1998–2006	0.7300	0.0000	0.7140	0.0040
7-day Eurodollar 1988–2006	0.0020	0.0000	0.0000	0.0000
7-day Eurodollar 1988–1997	0.0060	0.0060	0.0100	0.0140
7-day Eurodollar 1998–2006	0.0400	0.0440	0.1040	0.0520
Japanese yen 1988–2006	0.0780	0.0000	0.0200	0.0120
Japanese yen 1988–1997	0.0360	0.0760	0.1100	0.1440
Japanese yen 1998–2006	0.6048	0.1060	0.0360	0.1520

Notes: (i) $\hat{M}(m), m = 1,2,3,4$, are our proposed derivative tests, given in (4.1); (ii) The bootstrap p-values are calculated by the smoothed nonparametric transition density-based bootstrap procedure described in Section 5 with 500 bootstrap iterations.

bit surprisingly, the four derivative tests $\hat{M}(m)$, $m = 1,2,3,4$ all fail to reject the Markov hypothesis. However, for two subsamples, $\hat{M}(2)$ and $\hat{M}(4)$ reject the null at the 5% level, while $\hat{M}(1)$ and $\hat{M}(3)$ do not reject the null hypothesis. These results suggest that the violation of the Markov property may come from the conditional variance and kurtosis dynamics of S&P 500 returns. For 7-day Eurodollar rate changes, for both the whole sample and the first subsample, all four derivative tests firmly reject the null hypothesis at the 5% level. For the second subsample, $\hat{M}(1)$ and $\hat{M}(2)$ reject the null at the 5% level, but $\hat{M}(3)$ and $\hat{M}(4)$ do not. It seems that the violation of the Markov property for the 7-day Eurodollar rate comes from both mean and variance dynamics, and also possibly from higher-order moment dynamics. For Japanese yen changes, $\hat{M}(2)$, $\hat{M}(3)$, and $\hat{M}(4)$ tests strongly reject the null hypothesis at the 5% level, for the whole sample. However, the results from both subsamples are less clear. For the first subsample, only $\hat{M}(1)$ rejects the null hypothesis at the 5% level, and for the second subsample, only $\hat{M}(3)$ rejects the null hypothesis at the 5% level. To sum up, for all three financial series, we find strong evidence of violation of the Markov property in the conditional variance, among other things. This is consistent with the popular use of such non-Markovian models as generalized autoregressive conditional heteroskedasticity (GARCH) and stochastic volatility models in capturing the dynamics of price sequences in the literature.

As many financial time series have been documented to have a long memory property, which is non-Markov, we also apply Lobato and Robinson's (1998) test for the long memory property. Results (not reported here) show that there is no evidence of long memory for S&P 500 returns and Japanese yen returns in the whole sample and two subsamples, while there is some evidence of long memory for 7-day Eurodollar rate changes. Thus, we can not rule out the possibility that the rejection of the Markov property of 7-day Eurodollar rate may be due to the long memory property. Indeed, the evidence of 7-day Eurodollar rate changes against the Markov property is strongest among three time series.

The documented evidence against the Markov property casts some new thoughts on financial modeling. Although most popular stochastic differential equation models exhibit mathematical elegance and tractability, they may not be an adequate representation of the dynamics of the underlying process, due to the Markov assumption. Other modeling schemes, which allow for the non-Markov property, may be needed to better capture the dynamics of financial time series.

6. CONCLUSION

The Markov property is one of most fundamental properties in stochastic processes. Without justification, this property has been taken for granted in many economic and financial models, especially in continuous-time finance models. We propose a conditional characteristic function-based test for the Markov property in a spectral framework. The use of the conditional characteristic function, which is consistently estimated nonparametrically, allows us to check departures from the Markov property in all conditional moments, and the frequency domain approach, which checks many lags in a pairwise manner, provides a nice solution to tackling the difficulty of the curse of dimensionality associated with testing for the Markov property. To overcome the adverse impact of the first-stage nonparametric estimation of the conditional characteristic function, we use the smoothed nonparametric transition density-based bootstrap procedure, which provides reasonable sizes and powers for the proposed test in finite samples. We apply our test to three important financial time series and find some evidence that the Markov assumption may not be suitable for many financial time series.

NOTES

1. There are other existing tests for conditional independence of continuous variables in the literature. Linton and Gozalo (1997) propose two nonparamtric tests for conditional independence based on a generalization of the empirical distribution function. Su and White (2007, 2008) check conditional independence by the Hellinger distance and empirical characteristic function respectively. These tests can be used to test the Markov property. However, they are expected to encounter the "curse of dimensionality" problem because the Markov property implies that conditional independence must hold for an infinite number of lags.

2. Here we focus on the Markov property of order 1, which is the main interest of economic and financial modeling. However, our approach can be generalized to test the Markov property of order $p : P\left(\mathbf{X}_{t+1} \leq \mathbf{x}|\mathcal{I}_t\right) = P\left(\mathbf{X}_{t+1} \leq \mathbf{x}|\mathbf{X}_t, \mathbf{X}_{t-1}, ..., \mathbf{X}_{t-p+1}\right)$ for p fixed.

3. A multivariate Taylor series expansion can be obtained when $d > 1$. Since the expression is tedious, we do not present it here.

4. The extension is substantial since we use nonparametric estimation in the first stage and $\{Z_t(\mathbf{u})\}$ is not independent and identically distributed (i.i.d.) under $\mathbb{H}_0$.

5. See Masry (1996a, 1996b) for detailed explanations of these notations.

6. If $W(\mathbf{u})$ is differentiable, then this implies that its derivative $(\partial/\partial u_a)W(\mathbf{u})$ is an even function of u_a for $a = 1, ..., d$.

7. If $\mathbf{X}_t$ takes on discrete values, we can estimate $\varphi(\mathbf{u}|\mathbf{X}_t)$ via a frequency approach, namely replacing $\mathbf{K}_h(\mathbf{x}-\mathbf{X})$ with $\mathbf{1}(\mathbf{x}-\mathbf{X})$, where $\mathbf{1}(\cdot)$ is the indicator function. If $\mathbf{X}_t$ is a mix of discrete and continuous variables, e.g., $\mathbf{X}_t = (\mathbf{X}_t^d, \mathbf{X}_t^c)$, where $\mathbf{X}_t^d$ and $\mathbf{X}_t^c$ denote discrete and continuous components, respectively, following Li and Racine (2007), we can replace $\mathbf{K}_h(\cdot)$ with

$$W_\gamma(\mathbf{x},\mathbf{X}_s)=\mathbf{K}_h\left(\mathbf{x}^c-\mathbf{X}_s^c\right)\mathbf{L}_\lambda\left(\mathbf{x}^d,\mathbf{X}_s^d\right),$$

where $\gamma=(h,\lambda)$. And $\mathbf{L}_\lambda(\cdot)$ is the kernel function for the discrete components defined as

$$\mathbf{L}_\lambda\left(\mathbf{x}^d,\mathbf{X}_s^d\right)=\prod_{a=1}^{d}\lambda_a^{\mathbf{1}\left(X_{as}^d\neq x_a^d\right)},$$

where $0\leq\lambda_a\leq 1$ is the smoothing parameter for $\mathbf{X}_s^d$. Once we get a consistent estimator for $\varphi(\mathbf{u}|\mathbf{X}_t)$, we can calculate the generalized residual and construct the test statistic.

8. The proof strategy depends on Assumption 2. It seems plausible that one may relax Assumption 2 and rely on a more generous central limit theorem for degenerate U-statistics (e.g., Theorem 2.1 of Gao and Hong, 2008) but we may have to impose a more restrictive mixing condition as the cost. Due to its complexity, this will be left for our future research. On the other hand, Assumptions 1 and 2 do not imply each other. For example, consider a long memory process $\mathbf{X}_t=\sum_{j=0}^{\infty}\varphi_j\varepsilon_{t-j}$, where $\{\varepsilon_t\}\sim i.i.d.(0,1)$, $\varphi_j=\Gamma(j+d)/[\Gamma(d)\Gamma(j+1)]\approx\Gamma^{-1}(d)j^{d-1}$ as $j\to\infty$, where $\Gamma(\cdot)$ is the Gamma function. Define $\mathbf{X}_{qt}=\sum_{j=0}^{q}\varphi_j\varepsilon_{t-j}$, a q-dependent process. Then we have $\mathrm{E}\left(\mathbf{X}_t-\mathbf{X}_{qt}\right)^2\approx\Gamma^{-1}(d)q^{-1+2d}\cdot q^{-1}\sum_{j=q+1}^{\infty}(j/q)^{-2(1-d)}=O\left(q^{-1+2d}\right)$. Hence, Assumption 2 holds if $0<d\leq\frac{1}{4}$, but Assumption 1 is violated since $\{\mathbf{X}_t\}$ is not a strictly stationary β-mixing process.

9. Alternatively, we could impose Hansen's (2008) Assumption 3 on kernel functions, namely, for some $\Lambda<\infty$ and $L<\infty$, either $\mathbf{K}(\mathbf{u})=0$ for $\|\mathbf{u}\|>L$ and for all $\mathbf{u},\mathbf{u}'\in\mathbb{R}^d$, $\left|\mathbf{K}(\mathbf{u})-\mathbf{K}\left(\mathbf{u}'\right)\right|\leq\Lambda\left\|\mathbf{u}-\mathbf{u}'\right\|$; or $\mathbf{K}(\mathbf{u})$ is differentiable, $|(\partial/\partial\mathbf{u})\mathbf{K}(\mathbf{u})|\leq\Lambda$, and for some $\nu>1$, $|(\partial/\partial\mathbf{u})\mathbf{K}(\mathbf{u})|\leq\Lambda\|\mathbf{u}\|^{-\nu}$ for $\|\mathbf{u}\|>L$, where $\|\mathbf{u}\|\equiv\max(|u_1|,...,|u_d|)$. Here the kernel function is required to either have a truncated support and is Lipschitz or it has a bounded derivative with an integrable tail. Our proof could go through with this assumption, but the trade-off is a strengthening requirement on the bandwidth. Since the choice of the bandwidth is more important than the choice of the kernel, and many commonly used kernels have compact support, we only consider the case of the compact support of the kernel $\mathbf{K}(\mathbf{u})$ in our formal analysis. Nevertheless, we examine the effect of allowing kernels with support on $\mathbb{R}^d$ in our simulation study.

10. It is different from Paparoditis and Politis (2000), which requires different bandwidths. The reason why the same bandwidth works in our paper is that we use undersmoothing in the first stage, and the bias of the first-stage nonparametric estimation vanishes to 0 asymptotically. Therefore, we need not balance two bandwidths to obtain a good approximation of the asymptotic bias. This idea is shown in Theorem 2.1 (i) of Paparoditis and Politis (2000) in a different context.

11. Our simulation experiments show that results based on these two smoothed bootstrap procedures are very similar.

12. We thank Liangjun Su for providing the Matlab codes on computing Su and White's tests of conditional independence (2007, 2008).

13. We first generate 15 grid points $\mathbf{u}_0,\mathbf{v}_0$ from $N(0,1)$ and obtain $\mathbf{u}=[\mathbf{u}_0',-\mathbf{u}_0']'$ and $\mathbf{v}=[\mathbf{v}_0',-\mathbf{v}_0']'$ to ensure symmetry. Preliminary experiments with different numbers of grid points show that simulation results are not very sensitive to the choice of numbers. Concerned with the computational cost in the simulation study, we are satisfied with current results with 30 grid points.

14. We have tried the Parzen kernel for $k(\cdot)$, obtaining similar results (not reported here).

15. Following Ait-Sahalia (1997) and Amoro de Matos and Fernandes (2007), we use undersmoothing to ensure that the squared bias vanishes to zero faster than the variance. On the other hand, we have used the smoothed nonparametric conditional density bootstrap procedure, and hence the simulation results are expected not to be very sensitive to the choice of the bandwidth.

16. Ideally, the conditional distribution of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$ should be tested. Unfortunately, to our knowledge, no such test is available in the literature. Compared with some existing tests in the literature, Inoue's tests are model-free, allow for dependence in the data, and are robust against the heavy-tailed distributions observed in financial markets. Hence, they are most suitable here for preliminary testing.

REFERENCES

Aaronson, J., D. Gilat, & M. Keane (1992) On the structure of 1-dependent Markov chains. *Journal of Theoretical Probability* 5, 545–561.

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.

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

Ait-Sahalia, Y. (1997) Do Interest Rates Really Follow Continuous-Time Markov Diffusions? Working paper, Princeton University.

Ait-Sahalia, Y., J. Fan, and H. Peng (2009) Nonparametric transition-based tests for diffusions. *Journal the of American Statistical Association* 104, 1102–1116.

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

Amaro de Matos, J. & J. Rosario (2000) The Equilibrium Dynamics for an Endogenous Bid-Ask Spread in Competitive Financial Markets. Working paper, European University Institute and Universidade Nova de Lisboa.

Anderson, T. & J. Lund (1997) Estimating continuous time stochastic volatility models of the short term interest rate. *Journal of Econometrics* 77, 343–377.

Aviv, Y. & A. Pazgal (2005) A partially observed Markov decision process for dynamic pricing. *Management Science* 51, 1400–1416.

Bangia, A., F. Diebold, A. Kronimus, C. Schagen, & T. Schuermann (2002) Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance* 26, 445–474.

Bierens, H. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105–134.

Blume, L., D. Easley, & M. O'Hara (1994) Market statistics and technical analysis: The role of volume. *Journal of Finance* 49, 153–181.

Brown, B.M. (1971) Martingale limit theorems. *Annals of Mathematical Statistics* 42, 59–66.

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 (2009) Diagnosing Multivariate Continuous-Time Models with Application to Affine Term Structure Models. Working paper, Cornell University and University of Rochester.

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.

Darsow, W.F., B. Nguyen, & E.T. Olsen (1992) Copulas and Markov processes. *Illinois Journal of Mathematics* 36, 600–642.

Davies, R.B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.

Davies, R.B. (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.

Duan, J.C & K. Jacobs (2008) Is long memory necessary? An empirical investigation of nonnegative interest rate processes. *Journal of Empirical Finance* 15, 567–581.

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.

Easley, D. & M. O'Hara (1987) Price, trade size, and information in securities markets. *Journal of Financial Economics* 19, 69–90.

Easley, D. & M. O'Hara (1992) Time and the process of security price adjustment. *Journal of Finance* 47, 577–605.

Edwards, R. & J. Magee (1966) *Technical Analysis of Stock Trends.* John Magee.

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

Ericson, R. & A. Pakes (1995) Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies* 62(1), 53–82.

Fan, J. (1992) Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87, 998–1004.

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

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

Fan, Y. & Q. Li (1999) Root-N-consistent estimation of partially linear time series models. *Journal of Nonparametric Statistics* 11, 251–269.

Fan, Y., Q. Li, & I. Min (2006) A nonparametric bootstrap test of conditional distributions. *Econometric Theory* 22, 587–613.

Feller, W. (1959) Non-Markovian processes with the semi-group property. *Annals of Mathematical Statistics* 30, 1252–1253.

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., D. Hsieh, & G. Tauchen (1997) Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* 81, 159–192.

Gao, J. & Y. Hong (2008) Central limit theorems for generalized U-statistics with applications in nonparametric specification. *Journal of Nonparametric Statistics* 20, 61–76.

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

Hamilton, J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.

Hamilton J.D. & R. Susmel (1994) Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307–333.

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

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

Heath, D., R. Jarrow, & A. Morton (1992) Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica* 60, 77–105.

Hong, Y. (1999) Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association* 94, 1201–1220.

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.

Hong, Y. & H. White (2005) Asymptotic distribution theory for an entropy-based measure of serial dependence. *Econometrica* 73, 837–902.

Horowitz, J.L. (2003) Bootstrap methods for Markov processes. *Econometrica* 71, 1049–1082.

Ibragimov, R. (2007) Copula-Based Characterizations for Higher-Order Markov Processes. Working paper, Harvard University.

Inoue, A. (2001) Testing for distributional change in time series. *Econometric Theory* 17, 156–187.

Jarrow, R., D. Lando, & S. Turnbull (1997) A Markov model for the term structure of credit risk spreads. *Review of Financial Studies* 10, 481–523.

Jarrow, R. & S. Turnbull (1995) Pricing derivatives on financial securities subject to credit risk. *Journal of Finance* 50, 53–86.

Jiang, G. & J. Knight (1997) A nonparametric approach to the estimation of diffusion processes with an application to a short-term interest rate model. *Econometric Theory* 13, 615–645.

Kavvathas, D. (2001) Estimating Credit Rating Transition Probabilities for Corporate Bonds. Working paper, University of Chicago.

Kiefer, N.M. & C.E. Larson (2004) Testing Simple Markov Structures for Credit Rating Transitions. Working paper, Cornell University.

Kim, W. & O. Linton (2003) A Local Instrumental Variable Estimation Method for Generalized Additive Volatility Models. Working paper, Humboldt University of Berlin, London School of Economics.

Kydland, F.E. & E. Prescott (1982) Time to build and aggregate fluctuations. *Econometrica* 50, 1345–70.

Lando D. & T. Skϕdeberg (2002) Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance* 26, 423–444.

LeBaron, B. (1999) Technical trading rule profitability and foreign exchange intervention. *Journal of International Economics* 49, 125–143.

Lee, A.J. (1990) *U-Statistics: Theory and Practice*. Marcel Dekker.

Lévy, P. (1949) Exemple de processus pseudo-markoviens. *Comptes Rendus de l'Académie des Sciences* 228, 2004–2006.

Li, Q. & J.S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Linton, O. & P. Gozalo (1997) Conditional Independence Restrictions: Testing and Estimation. Working paper, Cowles Foundation for Research in Economics, Yale University.

Ljungqvist, L. & T.J. Sargent (2000) *Recursive Macroeconomic Theory*. MIT Press.

Lobato, I.N. & P.M. Robinson (1998) A nonparametric test for I(0). *Review of Economic Studies* 65, 475–495.

Loretan, M. & P.C.B Phillips (1994) Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of Empirical Finance* 1, 211–248.

Lucas, R. (1978) Asset prices in an exchange economy. *Econometrica* 46, 1429–45.

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, 659–81.

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. (1996a) Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 6, 571–599.

Masry, E. (1996b) Multivariate regression estimation local polynomial fitting for time series. *Stochastic Processes and Their Applications* 65, 81–101.

Masry, E. & J. Fan (1997) Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics* 24, 165–179.

Masry, E. & D. Tjøstheim (1997) Additive nonlinear ARX time series and projection estimates. *Econometric Theory* 13, 214–252.

Matús, F. (1996) On two-block-factor sequences and one-dependence. *Proceedings of the American Mathematical Society* 124, 1237–1242.

Matús, F. (1998) Combining m-dependence with Markovness. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques* 34, 407–423.

Mehra, R. & E. Prescott (1985) The equity premium: A puzzle. *Journal of Monetary Economics* 15, 145–61.

Mizutani, E. & S. Dreyfus (2004) Two stochastic dynamic programming problems by model-free actor-critic recurrent network learning in non-Markovian settings. *Proceedings of the IEEE-INNS International Joint Conference on Neural Networks.*

Pagan, A.R. & G.W. Schwert (1990) Testing for covariance stationarity in stock market data. *Economics Letters* 33, 165–70.

Paparoditis, E. & D.N. Politis (2000) The local bootstrap for kernel estimators under general dependence conditions. *Annals of the Institute of Statistical Mathematics* 52, 139–159.

Paparoditis, E. & D.N. Politis (2002) The local bootstrap for Markov processes. *Journal of Statistical Planning and Inference* 108, 301–328.

Platen, E. & R. Rebolledo (1996) Principles for modelling financial markets. *Journal of Applied Probability* 31, 601–613.

Romer, P. (1986) Increasing returns and long-run growth. *Journal of Political Economy* 5, 1002–1037.

Romer, P. (1990) Endogenous technological change. *Journal of Political Economy* 5, 71–102.

Rosenblatt, M. (1960) An aggregation problem for Markov chains. In R.E. Machol (ed.), *Information and Decision Processes*, pp. 87–92. McGraw-Hill.

Rosenblatt, M. & D. Slepian (1962) *N*th order Markov chains with every *N* variables independent. *Journal of the Society for Industrial and Applied Mathematics* 10, 537–549.

Ruppert, D. & M.P. Wand (1994) Multivariate weighted least squares regression. *Annals of Statistics* 22, 1346–1370.

Rust, J. (1994) Structural estimation of Markov decision processes. *Handbook of Econometrics* 4, 3081–3143.

Sargent, T. (1987) *Dynamic Macroeconomic Theory.* Harvard University Press.

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 (1993) Nonparametric tests of serial independence. In T. Subba Rao (ed.), *Developments in Time Series Analysis*: The Priestley Birthday Volume, pp. 207–229. Chapman & Hall.

Skaug, H.J. & D. Tjøstheim (1996) Measures of distance between densities with application to testing for serial independence. In P. Robinson & M. Rosenblatt (eds.), *Time Series Analysis in Memory of E. J. Hannan*, pp. 363–377. Springer.

Stinchcombe, M.B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.

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.

Su, L. & H. White (2008) Nonparametric Hellinger metric test for conditional independence. *Econometric Theory* 24, 829–864.

Uzawa, H. (1965) Optimum technical change in an aggregative model of economic growth. *International Economic Review* 6, 18–31.

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

Weintraub, G.Y., L.C. Benkard, & B. Van Roy (2008) Markov perfect industry dynamics with many firms. *Econometrica* 76, 1375–1411.

Yoshihara, K. (1976) Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrsch. Verw. Gebiete* 35, 237–252.

Zhu, X. (1992) Optimal fiscal policy in a stochastic growth model. *Journal of Economic Theory* 2, 250–289.

APPENDIX

Throughout the Appendix, we let $\tilde{M}$ be defined in the same way as $\hat{M}$ in (2.19) with $\hat{Z}_t(\mathbf{u})$ replaced by $Z_t(\mathbf{u})$. Also, $C \in (1, \infty)$ denotes a generic bounded constant.

Proof of Theorem 1. The proof of Theorem 1 consists of the proofs of Theorems A.1–A.3 below. ■

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

THEOREM A.2. *Let* $\tilde{M}_q$ *be defined as* $\tilde{M}$, *with* $\{\mathbf{X}_{q,t}\}_{t=1}^{T}$ *replacing* $\{\mathbf{X}_t\}_{t=1}^{T}$, $\{\varphi_{qt}(\mathbf{u})\}_{t=1}^{T}$ *replacing* $\{\varphi(\mathbf{u}|\mathbf{X}_{t-1})\}_{t=1}^{T}$, *where* $\{\mathbf{X}_{q,t}\}$ *and* $\{\varphi_{qt}(\mathbf{u})\}$ *are as in Assumption 2. Then under the conditions of Theorem 1 and* $q = p^{1+1/(4b-2)}(\ln^2 T)^{1/(2b-1)}$, $\tilde{M}_q - \tilde{M} \xrightarrow{p} 0$.

THEOREM A.3. *Under the conditions of Theorem 1 and* $q = p^{1+1/(4b-2)}(\ln^2 T)^{1/(2b-1)}$, $\tilde{M}_q \xrightarrow{d} N(0,1)$.

Proof of Theorem A.1. Put $T_j \equiv T - |j|$, and let $\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})$ be defined in the same way as $\hat{\Gamma}_j(\mathbf{u},\mathbf{v})$ in (2.11), with $\hat{Z}_t(\mathbf{u})$ replaced by $Z_t(\mathbf{u})$. To show $\hat{M} - \tilde{M} \xrightarrow{p} 0$, it suffices to show

$$\hat{D}^{-1/2} \iint \sum_{j=1}^{T-1} k^2(j/p) T_j [|\hat{\Gamma}_j(\mathbf{u},\mathbf{v})|^2 - |\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|^2] dW(\mathbf{u}) dW(\mathbf{v}) \xrightarrow{p} 0, \tag{A.1}$$

$p^{-1}(\hat{C} - \tilde{C}) = O_P(T^{-1/2})$, and $p^{-1}(\hat{D} - \tilde{D}) = o_P(1)$, where $\tilde{C}$ and $\tilde{D}$ are defined in the same way as $\hat{C}$ and $\hat{D}$ in (2.19), respectively, with $\hat{Z}_t(\mathbf{u})$ replaced by $Z_t(\mathbf{u})$. For space, we focus on the proof of (A.1); the proofs for $p^{-1}(\hat{C} - \tilde{C}) = O_P(T^{-1/2})$ and $p^{-1}(\hat{D} - \tilde{D}) = o_P(1)$ are straightforward. We note that it is necessary to obtain the convergence rate $O_P(pT^{-1/2})$ for $\hat{C} - \tilde{C}$ to ensure that replacing $\hat{C}$ with $\tilde{C}$ has asymptotically negligible impact given $p/T \to 0$.

To show (A.1), we first decompose

$$\iint \sum_{j=1}^{T-1} k^2(j/p) T_j \left[|\hat{\Gamma}_j(\mathbf{u},\mathbf{v})|^2 - |\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|^2\right] dW(\mathbf{u}) dW(\mathbf{v}) = \hat{A}_1 + 2\mathrm{Re}(\hat{A}_2), \tag{A.2}$$

where

$$\hat{A}_1 = \iint \sum_{j=1}^{T-1} k^2(j/p) T_j \left|\hat{\Gamma}_j(\mathbf{u},\mathbf{v}) - \tilde{\Gamma}_j(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u}) dW(\mathbf{v}),$$

$$\hat{A}_2 = \iint \sum_{j=1}^{T-1} k^2(j/p) T_j \left[\hat{\Gamma}_j(\mathbf{u},\mathbf{v}) - \tilde{\Gamma}_j(\mathbf{u},\mathbf{v})\right] \tilde{\Gamma}_j(\mathbf{u},\mathbf{v})^* dW(\mathbf{u}) dW(\mathbf{v}),$$

where $\mathrm{Re}(\hat{A}_2)$ is the real part of $\hat{A}_2$, and $\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})^*$ is the complex conjugate of $\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})$. Then (A.1) follows from Propositions A.1 and A.2 below, and $p \to \infty$ as $T \to \infty$. ■

PROPOSITION A.1. *Under the conditions of Theorem 1,* $p^{-1/2}\hat{A}_1 \xrightarrow{p} 0$.

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

Proof of Proposition A.1. Put $\psi_t(\mathbf{v}) \equiv e^{i\mathbf{v}'\mathbf{X}_t} - \varphi(\mathbf{v})$ and $\varphi(\mathbf{v}) \equiv \mathrm{E}(e^{i\mathbf{v}'\mathbf{X}_t})$. Then straightforward algebra yields that for $j > 0$,

$$
\begin{aligned}
&\hat{\Gamma}_j(\mathbf{u},\mathbf{v}) - \tilde{\Gamma}_j(\mathbf{u},\mathbf{v}) \\
&\quad = T_j^{-1} \sum_{t=j+1}^{T} \left[\varphi\left(\mathbf{u}|\mathbf{X}_{t-1}\right) - \hat{\varphi}\left(\mathbf{u}|\mathbf{X}_{t-1}\right)\right] \psi_{t-j}(\mathbf{v}) + \left[\varphi(\mathbf{v}) - \hat{\varphi}(\mathbf{v})\right] T_j^{-1} \\
&\qquad \sum_{t=j+1}^{T} \left[\varphi\left(\mathbf{u}|\mathbf{X}_{t-1}\right) - \hat{\varphi}\left(\mathbf{u}|\mathbf{X}_{t-1}\right)\right] \\
&\quad = \hat{B}_{1j}(\mathbf{u},\mathbf{v}) + \hat{B}_{2j}(\mathbf{u},\mathbf{v}), \quad \text{say.}
\end{aligned} \tag{A.3}
$$

It follows that $\hat{A}_1 \le 2\sum_{a=1}^{2} \sum_{j=1}^{T-1} k^2(j/p) T_j \iint |\hat{B}_{aj}(\mathbf{u},\mathbf{v})|^2 dW(\mathbf{u}) dW(\mathbf{v})$. Proposition A.1 follows from Lemmas A.1 and A.2 below. ■

Lemma A.1. $p^{-1/2} \sum_{j=1}^{T-1} k^2(j/p) T_j \iint |\hat{B}_{1j}(\mathbf{u},\mathbf{v})|^2 dW(\mathbf{u}) dW(\mathbf{v}) = o_P(1)$.

Lemma A.2. $p^{-1/2} \sum_{j=1}^{T-1} k^2(j/p) T_j \iint |\hat{B}_{2j}(\mathbf{u},\mathbf{v})|^2 dW(\mathbf{u}) dW(\mathbf{v}) = o_P(1)$.

We now show these lemmas. Throughout, we put $a_T(j) \equiv k^2(j/p) T_j^{-1}$.

Proof of Lemma A.1. We write

$$
\begin{aligned}
\hat{B}_{1j}(\mathbf{u},\mathbf{v}) &= T_j^{-1} \sum_{t=j+1}^{T} \left[\varphi(\mathbf{u}|\mathbf{X}_{t-1}) - \sum_{s=2}^{T} \hat{W}\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right) e^{i\mathbf{u}'\mathbf{X}_s}\right] \psi_{t-j}(\mathbf{v}) \\
&= -T_j^{-1} \sum_{t=j+1}^{T} \left[\sum_{s=2}^{T} \hat{W}\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right) \varphi(\mathbf{u}|\mathbf{X}_{s-1}) - \varphi(\mathbf{u}|\mathbf{X}_{t-1})\right] \psi_{t-j}(\mathbf{v}) \\
&\quad - T_j^{-1} \sum_{t=j+1}^{T} \sum_{s=2}^{T} \hat{W}\left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h}\right) Z_s(\mathbf{u}) \psi_{t-j}(\mathbf{v}) \\
&= -\hat{B}_{11j}(\mathbf{u},\mathbf{v}) - \hat{B}_{12j}(\mathbf{u},\mathbf{v}), \quad \text{say.}
\end{aligned} \tag{A.4}
$$

For the first term, we further decompose

$$
\begin{aligned}
&\hat{B}_{11j}(\mathbf{u},\mathbf{v}) \\
&\quad = T_j^{-1} \sum_{t=j+1}^{T} \mathbf{e}_1' h^{r+1} \mathbf{S}_T^{-1}(\mathbf{X}_{t-1}) \mathbf{B}_T(\mathbf{X}_{t-1}) \mathbf{D}_{r+1}(\mathbf{u},\mathbf{X}_{t-1}) \psi_{t-j}(\mathbf{v}) \\
&\qquad + T_j^{-1} \sum_{t=j+1}^{T} \mathbf{e}_1' \mathbf{S}_T^{-1}(\mathbf{X}_{t-1}) \mathbf{R}_T(\mathbf{u},\mathbf{X}_{t-1}) \psi_{t-j}(\mathbf{v}) \\
&\quad = \hat{B}_{111j}(\mathbf{u},\mathbf{v}) + \hat{B}_{112j}(\mathbf{u},\mathbf{v}), \quad \text{say,}
\end{aligned} \tag{A.5}
$$

where $\mathbf{B}_T(\mathbf{x})$ is an $N \times N_{r+1}$ matrix

$$\mathbf{B}_T(\mathbf{x}) = \begin{bmatrix} S_{0,r+1} \\ \vdots \\ S_{r,r+1} \end{bmatrix},$$

$S_{j,r+1}$ is of dimension $N_j \times N_{r+1}$, $\mathbf{D}_{r+1}(\mathbf{u},\mathbf{x})$ is obtained by arranging the N_{r+1} elements of the derivatives $1/\mathbf{j}!\varphi^{(\mathbf{j})}(\mathbf{u}|\mathbf{X}_{t-1}=\mathbf{x})$ for $|\mathbf{j}| = r+1$ as a column using the lexicographical order, $\mathbf{R}_T(\mathbf{u},\mathbf{x})$ is an $N \times 1$ vector

$$\mathbf{R}_T(\mathbf{u},\mathbf{x}) = \begin{bmatrix} \mathbf{R}_0 \\ \vdots \\ \mathbf{R}_r \end{bmatrix},$$

$\mathbf{R}_{|\mathbf{j}|}$ is of dimension $N_{|\mathbf{j}|} \times 1$, with its lth element $\left(\mathbf{R}_{|\mathbf{j}|}\right)_l = \gamma_{g_{|\mathbf{j}|}(l)}$, and

$$\gamma_{\mathbf{j}} = (r+1) \sum_{|\mathbf{l}|=r+1} \frac{h^{r+1}}{\mathbf{l}!(T-1)} \sum_{s=2}^{T} \left(\frac{\mathbf{X}_{s-1}-\mathbf{x}}{h}\right)^{\mathbf{l}+\mathbf{j}} \mathbf{K}_h(\mathbf{X}_{s-1}-\mathbf{x})$$
$$\times \int_0^1 \left[\varphi^{(\mathbf{l})}\left(\mathbf{u}|\mathbf{X}_{t-1}=\mathbf{x}+w\left(\mathbf{X}_{s-1}-\mathbf{x}\right)\right) - \varphi^{(\mathbf{l})}\left(\mathbf{u}|\mathbf{X}_{t-1}=\mathbf{x}\right)\right](1-w)^r\, dw. \quad \textbf{(A.6)}$$

For the first term in (A.5), we have

$$\begin{aligned}
\hat{B}_{111j}(\mathbf{u},\mathbf{v}) = \Big\{ & \mathrm{E}\left[\mathbf{e}_1' h^{r+1}\tilde{\mathbf{S}}(\mathbf{X}_{t-1})^{-1}\tilde{\mathbf{B}}(\mathbf{X}_{t-1})\mathbf{D}_{r+1}(\mathbf{u},\mathbf{X}_{t-1})\psi_{t-j}(\mathbf{v})\right] \\
& + T_j^{-1}\sum_{t=j+1}^{T}\Big\{\mathbf{e}_1' h^{r+1}\tilde{\mathbf{S}}(\mathbf{X}_{t-1})^{-1}\tilde{\mathbf{B}}(\mathbf{X}_{t-1}) \\
& \qquad \times \mathbf{D}_{r+1}(\mathbf{u},\mathbf{X}_{t-1})\psi_{t-j}(\mathbf{v}) \\
& \qquad - \mathrm{E}\Big[\mathbf{e}_1' h^{r+1}\tilde{\mathbf{S}}(\mathbf{X}_{t-1})^{-1}\tilde{\mathbf{B}}(\mathbf{X}_{t-1}) \\
& \qquad \times \mathbf{D}_{r+1}(\mathbf{u},\mathbf{X}_{t-1})\psi_{t-j}(\mathbf{v})\Big]\Big\}\Big\}[1+o_P(1)] \\
= & \left[\hat{B}_{1111j}(\mathbf{u},\mathbf{v}) + \hat{B}_{1112j}(\mathbf{u},\mathbf{v})\right][1+o_P(1)], \quad \text{say}, \qquad \textbf{(A.7)}
\end{aligned}$$

where $\tilde{\mathbf{S}}(\mathbf{x}) \equiv \mathrm{E}[\mathbf{S}_T(\mathbf{x})]$ and $\tilde{\mathbf{B}}(\mathbf{x}) \equiv \mathrm{E}[\mathbf{B}_T(\mathbf{x})]$.

For the first term in (A.7), we have

$$\begin{aligned}
& \iint \left|\hat{B}_{1111j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v}) \\
& \le Ch^{2(r+1)} \iint \Big|\beta(j)\Big\|\mathbf{e}_1'\tilde{\mathbf{S}}(\mathbf{X}_{t-1})^{-1}\tilde{\mathbf{B}}(\mathbf{X}_{t-1}) \\
& \quad \times \mathbf{D}_{r+1}(\mathbf{u},\mathbf{X}_{t-1})\Big\|_\infty \left\|\psi_{t-j}(\mathbf{v})\right\|_\infty\Big|^2 dW(\mathbf{u})\,dW(\mathbf{v}) \\
& \le C\beta^2(j)h^{2(r+1)},
\end{aligned}$$

where we have used the mixing inequality and Assumption 1. It follows that

$$p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)\,T_j \iint \left|\hat{B}_{1111j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v}) = o_P(1), \tag{A.8}$$

where we have used the fact that

$$\sum_{j=1}^{T-1} a_T(j) = \sum_{j=1}^{T-1} k^2(j/p)T_j^{-1} = O(p/T). \tag{A.9}$$

For the second term in (A.7), we have

$$\begin{aligned}
&\iint \mathrm{E}\left|\hat{B}_{1112j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v})\\
&\quad = 2T_j^{-2}\sum_{\tau=j+1}^{T}\sum_{t=\tau+1}^{T}\iint \mathrm{E}\Big[\mathbf{e}_1' h^{r+1}\tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1}\tilde{\mathbf{B}}\left(\mathbf{X}_{t-1}\right)\\
&\qquad \times \mathbf{D}_{r+1}\left(\mathbf{u},\mathbf{X}_{t-1}\right)\psi_{t-j}(\mathbf{v}) - \hat{B}_{1111j}(\mathbf{u},\mathbf{v})\Big]\Big[\mathbf{e}_1' h^{r+1}\tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1}\tilde{\mathbf{B}}\left(\mathbf{X}_{t-1}\right)\\
&\qquad \times \mathbf{D}_{r+1}\left(\mathbf{u},\mathbf{X}_{\tau-1}\right)\psi_{\tau-j}(\mathbf{v}) - \hat{B}_{1111j}(\mathbf{u},\mathbf{v})\Big]^{*} dW(\mathbf{u})\,dW(\mathbf{v})\\
&\qquad + T_j^{-1}\iint \mathrm{E}\Big|\mathbf{e}_1' h^{r+1}\tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1}\tilde{\mathbf{B}}\left(\mathbf{X}_{t-1}\right)\\
&\qquad \times \mathbf{D}_{r+1}\left(\mathbf{u},\mathbf{X}_{t-1}\right)\psi_{t-j}(\mathbf{v})\Big|^2 dW(\mathbf{u})dW(\mathbf{v})\\
&\quad \le CT_j^{-1}\sum_{l=1}^{T-j}\left(1-\frac{l}{T_j}\right)\beta(l)\,h^{2(r+1)} + CT_j^{-1}h^{2(r+1)} \le CT_j^{-1}h^{2(r+1)},
\end{aligned}$$

where we have used Assumption 1 and the mixing inequality. It follows from (A.9) and Chebychev's inequality that

$$p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)\,T_j \iint \left|\hat{B}_{1112j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v}) = o_P(1). \tag{A.10}$$

By Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)\,T_j \iint \hat{B}_{1111j}(\mathbf{u},\mathbf{v})\,\hat{B}_{1112j}^{*}(\mathbf{u},\mathbf{v})\,dW(\mathbf{u})\,dW(\mathbf{v})\\
&\quad \le p^{-1/2}\left[\sum_{j=1}^{T-1} k^2(j/p)\,T_j \iint \left|\hat{B}_{1111j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v})\right]^{1/2}\\
&\qquad \left[\sum_{j=1}^{T-1} k^2(j/p)\,T_j \iint \left|\hat{B}_{1112j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v})\right]^{1/2}\\
&\quad = o_P(1).
\end{aligned} \tag{A.11}$$

Combining (A.8), (A.10), and (A.11), we obtain

$$
\begin{aligned}
&p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)\,T_j \iint \left|\hat{B}_{111j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v}) \\
&\quad = p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)\,T_j \iint \left|\hat{B}_{1111j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v}) \\
&\qquad + p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)\,T_j \iint \left|\hat{B}_{1112j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v}) \\
&\qquad - 2p^{-1/2}\mathrm{Re}\sum_{j=1}^{T-1} k^2(j/p)\,T_j \iint \hat{B}_{1111j}(\mathbf{u},\mathbf{v})\,\hat{B}^*_{1112j}(\mathbf{u},\mathbf{v})\,dW(\mathbf{u})\,dW(\mathbf{v}) \\
&\quad = o_P(1).
\end{aligned}
\tag{A.12}
$$

For the second term in (A.5), we have

$$
\begin{aligned}
\hat{B}_{112j}(\mathbf{u},\mathbf{v}) &= \Big\{\mathrm{E}\left[\mathbf{e}_1'\tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1}\tilde{\mathbf{R}}\left(\mathbf{u},\mathbf{X}_{t-1}\right)\psi_{t-j}(\mathbf{v})\right] \\
&\qquad + T_j^{-1}\sum_{t=j+1}^{T}\Big\{\mathbf{e}_1'\tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1}\tilde{\mathbf{R}}\left(\mathbf{u},\mathbf{X}_{t-1}\right)\psi_{t-j}(\mathbf{v}) \\
&\qquad\qquad - \mathrm{E}\left[\mathbf{e}_1'\tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1}\tilde{\mathbf{R}}\left(\mathbf{u},\mathbf{X}_{t-1}\right)\psi_{t-j}(\mathbf{v})\right]\Big\}\Big\}\left[1+o_P(1)\right] \\
&= \left[\hat{B}_{1121j}(\mathbf{u},\mathbf{v}) + \hat{B}_{1122j}(\mathbf{u},\mathbf{v})\right]\left[1+o_P(1)\right], \quad \text{say,}
\end{aligned}
\tag{A.13}
$$

where $\tilde{\mathbf{R}}(\mathbf{x})$ is an $N\times 1$ vector

$$
\tilde{\mathbf{R}}(\mathbf{u},\mathbf{x}) = \begin{bmatrix}\tilde{\mathbf{R}}_0\\ \vdots \\ \tilde{\mathbf{R}}_r\end{bmatrix},
$$

$\tilde{\mathbf{R}}_{|\mathbf{j}|}$ is of dimension $N_{|\mathbf{j}|}\times 1$, with its lth element $\left(\tilde{\mathbf{R}}_{|\mathbf{j}|}\right)_l = \tilde{\gamma}_{g_{|\mathbf{j}|}(l)} \equiv \mathrm{E}\left[\gamma_{g_{|\mathbf{j}|}(l)}\right]$.

For the first term in (A.13), we have

$$
\begin{aligned}
&\iint \left|\hat{B}_{1121j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v}) \\
&\quad \le C\iint \beta^2(j)\left|\left\|\mathbf{e}_1'\tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1}\tilde{\mathbf{R}}\left(\mathbf{u},\mathbf{X}_{t-1}\right)\right\|_\infty \left\|\psi_{t-j}(\mathbf{v})\right\|_\infty\right|^2 dW(\mathbf{u})\,dW(\mathbf{v}) \\
&\quad \le C\beta^2(j)\,h^{2(r+1)},
\end{aligned}
$$

where we have used the mixing inequality, Assumption 1, and the fact that

$$
\sup_{\mathbf{x}\in\mathbf{G}}\left|\tilde{\gamma}_{\mathbf{j}}\right|^2 = O_P\left(h^{2(r+1)}\right).
\tag{A.14}
$$

It follows that

$$p^{-\frac{1}{2}} \sum_{j=1}^{T-1} k^2 (j/p) T_j \iint \left| \hat{B}_{1121j} (\mathbf{u}, \mathbf{v}) \right|^2 dW (\mathbf{u}) dW (\mathbf{v}) = o_P (1), \quad \textbf{(A.15)}$$

where we have used (A.9).

For the second term in (A.13), we have

$$\begin{aligned}
&\iint \mathrm{E} \left| \hat{B}_{1122j} (\mathbf{u}, \mathbf{v}) \right|^2 dW (\mathbf{u}) dW (\mathbf{v}) \\
&= 2T_j^{-2} \sum_{\tau=j+1}^{T} \sum_{t=\tau+1}^{T} \iint \mathrm{E} \left[\mathbf{e}_1' \tilde{\mathbf{S}} (\mathbf{X}_{t-1})^{-1} \tilde{\mathbf{R}} (\mathbf{u}, \mathbf{X}_{t-1}) \psi_{t-j} (\mathbf{v}) - \hat{B}_{1121j} (\mathbf{u}, \mathbf{v}) \right] \\
&\quad \times \left[\mathbf{e}_1' \tilde{\mathbf{S}} (\mathbf{X}_{t-1})^{-1} \tilde{\mathbf{R}} (\mathbf{u}, \mathbf{X}_{t-1}) \psi_{\tau-j} (\mathbf{v}) - \hat{B}_{1121j} (\mathbf{u}, \mathbf{v}) \right]^* dW (\mathbf{u}) dW (\mathbf{v}) \\
&\quad + T_j^{-1} \iint \mathrm{E} \left| \mathbf{e}_1' \tilde{\mathbf{S}} (\mathbf{X}_{t-1})^{-1} \tilde{\mathbf{R}} (\mathbf{u}, \mathbf{X}_{t-1}) \psi_{t-j} (\mathbf{v}) \right|^2 dW (\mathbf{u}) dW (\mathbf{v}) \\
&\le CT_j^{-1} \sum_{l=1}^{T-j} \left(1 - \frac{l}{T_j} \right) \beta (l) h^{2(r+1)} + CT_j^{-1} h^{2(r+1)} \le CT_j^{-1} h^{2(r+1)},
\end{aligned}$$

where we have used (A.14), Assumption 1, and the mixing inequality. It follows from (A.9) and Chebychev's inequality that

$$p^{-1/2} \sum_{j=1}^{T-1} k^2 (j/p) T_j \iint \left| \hat{B}_{1122j} (\mathbf{u}, \mathbf{v}) \right|^2 dW (\mathbf{u}) dW (\mathbf{v}) = o_P (1). \quad \textbf{(A.16)}$$

Combining (A.15) and (A.16), we obtain

$$p^{-1/2} \sum_{j=1}^{T-1} k^2 (j/p) T_j \iint \left| \hat{B}_{112j} (\mathbf{u}, \mathbf{v}) \right|^2 dW (\mathbf{u}) dW (\mathbf{v}) = o_P (1). \quad \textbf{(A.17)}$$

For the second term in (A.4), we have

$$\begin{aligned}
\hat{B}_{12j} (\mathbf{u}, \mathbf{v}) &= T_j^{-1} (T-1)^{-1} h^{-d} \sum_{t=j+1}^{T} \sum_{s=2}^{T\prime} \mathbf{e}_1 \tilde{\mathbf{S}} (\mathbf{X}_{t-1})^{-1} \Xi \\
&\quad \times \left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h} \right) Z_s (\mathbf{u}) \psi_{t-j} (\mathbf{v}) [1 + o_P (1)] \\
&= \Bigg\{ T_j^{-1} (T-1)^{-1} h^{-d} \sum_{t=2}^{T} \sum_{s=2}^{T} \mathbf{e}_1' \tilde{\mathbf{S}} (\mathbf{X}_{t-1})^{-1} \Xi \\
&\quad \times \left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h} \right) Z_s (\mathbf{u}) \psi_{t-j} (\mathbf{v}) \\
&\quad - T_j^{-1} (T-1)^{-1} h^{-d} \sum_{t=2}^{j} \sum_{s=2}^{T} \mathbf{e}_1' \tilde{\mathbf{S}} (\mathbf{X}_{t-1})^{-1} \Xi
\end{aligned}$$

$$\times \left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h} \right) Z_s(\mathbf{u}) \psi_{t-j}(\mathbf{v}) \Big\} [1 + o_P(1)]$$

$$= \left[\hat{B}_{121j}(\mathbf{u}, \mathbf{v}) - \hat{B}_{122j}(\mathbf{u}, \mathbf{v}) \right] [1 + o_P(1)], \tag{A.18}$$

where $\Xi(\mathbf{z}) \equiv \Theta(\mathbf{z}) \mathbf{K}(\mathbf{z})$. Now, introducing

$$\Phi_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{js} \right) = \mathbf{e}_1' \tilde{\mathbf{S}} \left(\mathbf{X}_{t-1} \right)^{-1} \Xi \left(\frac{\mathbf{X}_{s-1} - \mathbf{X}_{t-1}}{h} \right) Z_s(\mathbf{u}) \psi_{t-j}(\mathbf{v})$$
$$+ \mathbf{e}_1' \tilde{\mathbf{S}} \left(\mathbf{X}_{s-1} \right)^{-1} \Xi \left(\frac{\mathbf{X}_{t-1} - \mathbf{X}_{s-1}}{h} \right) Z_t(\mathbf{u}) \psi_{s-j}(\mathbf{v}),$$

where $\mathbf{Y}_{jt} = \left(\mathbf{X}_t, \mathbf{X}_{t-1}, \mathbf{X}_{t-j} \right)$, we can write $\hat{B}_{121j}(\mathbf{u}, \mathbf{v})$ as a U-statistic,

$$\hat{B}_{121j}(\mathbf{u}, \mathbf{v}) = \frac{1}{2} T_j^{-1} (T-1)^{-1} h^{-d} \sum_{t \neq s} \sum \Phi_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{js} \right)$$
$$+ \frac{1}{2} T_j^{-1} (T-1)^{-1} h^{-d} \sum_{t=2}^{T} \Phi_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{jt} \right).$$

For notational simplicity, we have suppressed the dependence on $\mathbf{u}$ and $\mathbf{v}$ of $\phi_j(Y_{jt}, Y_{js})$.

For the second term, it is easy to see that

$$\sum_{j=1}^{T-1} k^2(j/p) T_j \iint \left| \frac{1}{2} T_j^{-1} (T-1)^{-1} h^{-d} \sum_{t=2}^{T} \Phi_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{jt} \right) \right|^2 dW(\mathbf{u}) dW(\mathbf{v}) = o_P(1).$$

For the first term, we have

$$\frac{1}{2} T_j^{-1} (T-1)^{-1} h^{-d} \sum_{t \neq s} \sum \Phi_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{js} \right)$$
$$= T_j^{-1} h^{-d} \sum_{t=2}^{T} \Phi_{1j} \left(\mathbf{X}_{t-1} \right) + \frac{1}{2} T_j^{-1} (T-1)^{-1} h^{-d} \sum_{t > s} \sum \tilde{\Phi}_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{js} \right)$$
$$= T_j^{-1} h^{-d} \sum_{t=2}^{T} \Phi_{1j} \left(\mathbf{X}_{t-1} \right) + \tilde{\Phi}_j, \quad \text{say}, \tag{A.19}$$

where $\Phi_{1j}(\mathbf{y}) = \int \Phi_j \left(\mathbf{y}, \mathbf{Y}_{js} \right) dF_j \left(\mathbf{Y}_{js} \right) = \int \mathbf{e}_1' \tilde{\mathbf{S}} \left(\mathbf{X}_{s-1} \right)^{-1} \Xi \left(\frac{\mathbf{X}_{t-1} - \mathbf{X}_{s-1}}{h} \right) Z_t(\mathbf{u})$ $\psi_{s-j}(\mathbf{v}) dF_j \left(\mathbf{Y}_{js} \right)$ and $\tilde{\Phi}_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{js} \right) = \Phi_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{js} \right) - \Phi_{1j} \left(\mathbf{Y}_{jt} \right) - \Phi_{1j} \left(\mathbf{Y}_{js} \right)$.

Note we have made use of the fact that

$$\Phi_{0j} = \int \Phi_{1j} \left(\mathbf{Y}_{jt} \right) dF \left(\mathbf{Y}_{jt} \right)$$
$$= \iint \mathbf{e}_1' \tilde{\mathbf{S}} \left(\mathbf{X}_{s-1} \right)^{-1} \Xi \left(\frac{\mathbf{X}_{t-1} - \mathbf{X}_{s-1}}{h} \right) Z_t(\mathbf{u}) \psi_{s-j}(\mathbf{v}) dF_j \left(\mathbf{Y}_{js} \right) dF_j \left(\mathbf{Y}_{jt} \right) = 0.$$

For the first term in (A.19), we have

$$\begin{aligned}
&\operatorname{var}\left[T_j^{-1}h^{-d}\sum_{t=2}^{T}\Phi_{1j}\left(\mathbf{Y}_{jt}\right)\right]\\
&\quad= 2T_j^{-2}h^{-2d}\sum_{t>\tau}\sum\operatorname{cov}\left[\Phi_{1j}\left(\mathbf{Y}_{j,t-1}\right),\Phi_{1j}^{*}\left(\mathbf{Y}_{j,\tau-1}\right)\right]\\
&\qquad+T_j^{-2}(T-1)h^{-2d}\operatorname{var}\left[\Phi_{1j}\left(\mathbf{X}_{t-1}\right)\right]\\
&\quad\le CT_j^{-2}(T-1)h^{-2d}\sum_{l=1}^{T-1}\left(1-\frac{l}{T}\right)\beta(l)^{1-(1/\gamma)}\left[\mathrm{E}\left|\Phi_{1j}\left(\mathbf{X}_{t-1}\right)\right|^{2\gamma}\right]^{1/\gamma}\\
&\qquad+T_j^{-2}(T-1)h^{-2d}\operatorname{var}\left[\Phi_{1j}\left(\mathbf{X}_{t-1}\right)\right]\\
&\quad\le CT_j^{-2}(T-1)h^{-2d}\left[\mathrm{E}\left|\Phi_{1j}\left(\mathbf{X}_{t-1}\right)\right|^{2\gamma}\right]^{1/\gamma}\\
&\qquad+T_j^{-2}(T-1)h^{-2d}\operatorname{var}\left[\Phi_{1j}\left(\mathbf{X}_{t-1}\right)\right],
\end{aligned}$$

where $\gamma=\frac{\nu}{\nu-1}+\varepsilon$ and $\varepsilon>0$. Note that we have used Assumption 1 and the mixing inequality. Put $D=h^{-d/\nu}$, where ν is defined in Assumption 1. It follows that

$$\begin{aligned}
&p^{-1/2}\sum_{j=1}^{D}k^2(j/p)T_j\iint\mathrm{E}\left|T_j^{-1}h^{-d}\sum_{t=2}^{T}\Phi_{1j}\left(\mathbf{X}_{t-1}\right)\right|^2dW(\mathbf{u})\,dW(\mathbf{v})\\
&\quad\le CDp^{-1/2}=o(1),
\end{aligned}\tag{A.20}$$

where we have used the fact that $h^{-d}\Phi_{1j}(\mathbf{z})$ is bounded in probability. At the same time, we have

$$\begin{aligned}
&p^{-1/2}\sum_{j=D+1}^{T-1}k^2(j/p)T_j\iint\mathrm{E}\left|T_j^{-1}h^{-d}\sum_{t=2}^{T}\Phi_{1j}\left(\mathbf{X}_{t-1}\right)\right|^2dW(\mathbf{u})\,dW(\mathbf{v})\\
&\quad\le Cp^{-1/2}\sum_{j=D+1}^{T-1}k^2(j/p)\beta^2(j-1)h^{-2d}=O\left(h^{-d/\nu}p^{-1/2}\right)=o(1),
\end{aligned}\tag{A.21}$$

where we have used Assumptions 1, and 4, and the fact $\left|\Phi_{1j}(\mathbf{x})\right|\le\beta(j-1)$ given the mixing inequality.

It follows from (A.20), (A.21), and Chebychev's inequality that

$$p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|T_j^{-1}\sum_{t=2}^{T}\Phi_{1j}\left(\mathbf{X}_{t-1}\right)\right|^2dW(\mathbf{u})\,dW(\mathbf{v})=o_P(1).\tag{A.22}$$

For the second term in (A.19), we have

$$\mathrm{E}(\tilde{\Phi}_j^2)=T_j^{-2}(T-1)^{-2}h^{-2d}\sum_{t\neq s}\sum_{t'\neq s'}\mathrm{E}\left[\tilde{\Phi}_j\left(\mathbf{Y}_{jt},\mathbf{Y}_{js}\right)\tilde{\Phi}_j^{*}\left(\mathbf{Y}_{jt'},\mathbf{Y}_{js'}\right)\right].$$

Following Yoshihara (1976) and Lee (1990), we split it into two types: (a) those for which t, s, t', s' are all distinct; and (b) those remaining.

For terms of type (a), we have

$$\begin{aligned}
&\left| \sum_{t,s,t',s'} \mathrm{E}\left[\tilde{\Phi}_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{js} \right) \tilde{\Phi}_j^* \left(\mathbf{Y}_{jt'}, \mathbf{Y}_{js'} \right) \right] \right| \\
&\quad \leq 4! \sum_{t=2}^{T} \sum_{s=2}^{T-1} \sum_{t'=2}^{T-1} \sum_{s'=2}^{T-1} \left| \mathrm{E}\left[\tilde{\Phi}_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{jt+s} \right) \tilde{\Phi}_j^* \left(\mathbf{Y}_{jt+s+t'}, \mathbf{Y}_{jt+s+t'+s'} \right) \right] \right| \\
&\quad \leq 4! \sum_{t=2}^{T} \sum_{2 \leq t', s' \leq s} \left| \mathrm{E}\left[\tilde{\Phi}_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{jt+s} \right) \tilde{\Phi}_j^* \left(\mathbf{Y}_{jt+s+t'}, \mathbf{Y}_{jt+s+t'+s'} \right) \right] \right| \\
&\qquad + 4! \sum_{t=2}^{T} \sum_{2 \leq s, s' \leq t'} \left| \mathrm{E}\left[\tilde{\Phi}_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{jt+s} \right) \tilde{\Phi}_j^* \left(\mathbf{Y}_{jt+s+t'}, \mathbf{Y}_{jt+s+t'+s'} \right) \right] \right| \\
&\qquad + 4! \sum_{t=2}^{T} \sum_{2 \leq s, t' \leq s'} \left| \mathrm{E}\left[\tilde{\Phi}_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{jt+s} \right) \tilde{\Phi}_j^* \left(\mathbf{Y}_{jt+s+t'}, \mathbf{Y}_{jt+s+t'+s'} \right) \right] \right|. \qquad \textbf{(A.23)}
\end{aligned}$$

For the first term in (A.23), we have

$$\begin{aligned}
&\sum_{t=2}^{T} \sum_{2 \leq t', s' \leq s} \left| \mathrm{E}\left[\tilde{\Phi}_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{jt+s} \right) \tilde{\Phi}_j^* \left(\mathbf{Y}_{jt+s+t'}, \mathbf{Y}_{jt+s+t'+s'} \right) \right] \right| \\
&\quad \leq \sum_{t=2}^{T} \sum_{s=2}^{T} (s+1)^2 \beta^{\alpha/\alpha+1}(s) h^{2d/(\alpha+1)} \leq C(T-1) h^{2d/(\alpha+1)},
\end{aligned}$$

where $1 > \alpha > \frac{3}{\nu-3}$ and we have used Lemma 2 of Yoshihara (1976) and Assumption 1. For the second term in (A.23), we have

$$\begin{aligned}
&\sum_{t=2}^{T} \sum_{2 \leq s, s' \leq t'} \left| \mathrm{E}\left[\tilde{\Phi}_j \left(\mathbf{Y}_{jt}, \mathbf{Y}_{jt+s} \right) \tilde{\Phi}_j^* \left(\mathbf{Y}_{jt+s+t'}, \mathbf{Y}_{jt+s+t'+s'} \right) \right] \right| \\
&\quad \leq \sum_{t=2}^{T} \sum_{t'=2}^{T} \left(t'+1 \right)^2 \beta^{\alpha/\alpha+1} \left(t' \right) h^{2d/(\alpha+1)} \\
&\qquad + \sum_{t=2}^{T} \sum_{2 \leq s, s' \leq t'} \beta^{\alpha/\alpha+1}(s) h^{d/(\alpha+1)} \beta^{\alpha/\alpha+1} \left(s' \right) h^{d/(\alpha+1)} \\
&\quad \leq C(T-1)^2 h^{2d/(\alpha+1)},
\end{aligned}$$

where $1 > \alpha > \frac{3}{\nu-3}$ and we have used Lemma 2 of Yoshihara and Assumption 1. The third term is similar to the first term.

For terms of type (b), we also consider one case:

$$\left|\sum_{2\le t<t'\le T}\sum_{s=2}^{T}\mathrm{E}\left[\tilde{\Phi}_j\left(\mathbf{Y}_{jt},\mathbf{Y}_{js}\right)\tilde{\Phi}_j^*\left(\mathbf{Y}_{jt'},\mathbf{Y}_{js}\right)\right]\right| \le (T-1)^2 h^d\left[1+\sum_{j=1}^{T}\beta^{\alpha/\alpha+1}(j)\right]$$
$$\le C(T-1)^2 h^d,$$

where $1>\alpha>\frac{1}{\nu-1}$ and we have used Lemma 2 of Yoshihara (1976) and Assumption 1. For other cases, similar arguments apply.

Hence we have

$$p^{-\frac{1}{2}}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\tilde{\Phi}_j\right|^2 dW(\mathbf{u})\,dW(\mathbf{v})=o_P(1), \quad \textbf{(A.24)}$$

where we have used Chebychev's inequality and (A.9).

It follows from (A.19), (A.22), and (A.24) that

$$p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{121j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v})=o_P(1). \quad \textbf{(A.25)}$$

For the second term in (A.18), we have

$$\begin{aligned}
&\mathrm{E}\left|\hat{B}_{122j}(\mathbf{u},\mathbf{v})\right|^2\\
&\quad=2T_j^{-2}(T-1)^{-2}h^{-2d}\sum_{t=2}^{j}\sum_{t'=t+1}^{j}\sum_{s=2}^{T}\sum_{s'=s+1}^{T}\mathrm{E}g\left(\mathbf{X}_{t-1}\right)^{-1}\Xi\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\\
&\quad\quad\times Z_s(\mathbf{u})\,\psi_{t-j}(\mathbf{v})\,\mathbf{e}_1'\tilde{\mathbf{S}}\left(\mathbf{X}_{t'-1}\right)^{-1}\Xi\left(\frac{\mathbf{X}_{s'-1}-\mathbf{X}_{t'-1}}{h}\right)Z_{s'}(\mathbf{u})\,\psi_{t'-j}(\mathbf{v})\\
&\quad\quad+T_j^{-2}(T-1)^{-2}h^{-2d}\sum_{t=2}^{j}\sum_{s=2}^{T}\\
&\quad\quad\times\mathrm{E}\left|\mathbf{e}_1'\tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1}\Xi\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)Z_s(\mathbf{u})\,\psi_{t-j}(\mathbf{v})\right|^2\\
&\quad\le CT_j^{-2}j^2+CT_j^{-2}(T-1)^{-1}jh^{-d},
\end{aligned}$$

where we have used Assumptions 1 and 3.

It follows from (A.9) and Chebychev's inequality that

$$p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{122j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v})=o_P(1). \quad \textbf{(A.26)}$$

Then it follows from (A.4), (A.5), (A.12), (A.17), (A.25), and (A.26) that

$$p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{1j}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\,dW(\mathbf{v})=o_P(1). \quad \textbf{(A.27)}$$

The desired result of Lemma A.1 follows. ■

Proof of Lemma A.2. We write

$$
\begin{aligned}
&\hat{B}_{2j}(\mathbf{u},\mathbf{v}) \\
&= \left[\varphi(\mathbf{v})-\hat{\varphi}(\mathbf{v})\right] T_j^{-1} \sum_{t=j+1}^{T} \left[\varphi\left(\mathbf{u}|\mathbf{X}_{t-1}\right) - \sum_{s=2}^{T} \hat{W}\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) e^{i\mathbf{u}'\mathbf{X}_s}\right] \\
&= -\left[\varphi(\mathbf{v})-\hat{\varphi}(\mathbf{v})\right] T_j^{-1} \sum_{t=j+1}^{T} \left[\sum_{s=2}^{T} \hat{W}\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right)\varphi\left(\mathbf{u}|\mathbf{X}_{s-1}\right) - \varphi\left(\mathbf{u}|\mathbf{X}_{t-1}\right)\right] \\
&\quad - \left[\varphi(\mathbf{v})-\hat{\varphi}(\mathbf{v})\right] T_j^{-1} \sum_{t=j+1}^{T} \sum_{s=2}^{T} \hat{W}\left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) Z_s(\mathbf{u}) \\
&= -\hat{B}_{21j}(\mathbf{u},\mathbf{v}) - \hat{B}_{22j}(\mathbf{u},\mathbf{v}), \quad \text{say.} \qquad \textbf{(A.28)}
\end{aligned}
$$

We further decompose

$$
\begin{aligned}
\hat{B}_{21j}(\mathbf{u},\mathbf{v}) &= \left[\varphi(\mathbf{v})-\hat{\varphi}(\mathbf{v})\right] T_j^{-1} \sum_{t=j+1}^{T} \mathbf{e}_1' h^{r+1} \mathbf{S}_T^{-1}\left(\mathbf{X}_{t-1}\right) \mathbf{B}_T\left(\mathbf{X}_{t-1}\right) \mathbf{D}_{r+1}\left(\mathbf{X}_{t-1}\right) \\
&\quad + \left[\varphi(\mathbf{v})-\hat{\varphi}(\mathbf{v})\right] T_j^{-1} \sum_{t=j+1}^{T} \mathbf{e}_1' \mathbf{S}_T^{-1}\left(\mathbf{X}_{t-1}\right) \mathbf{R}_T\left(\mathbf{X}_{t-1}\right) \\
&= \hat{B}_{211j}(\mathbf{u},\mathbf{v}) + \hat{B}_{212j}(\mathbf{u},\mathbf{v}), \quad \text{say,} \qquad \textbf{(A.29)}
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{B}_{22j}(\mathbf{u},\mathbf{v}) &= \Bigg\{ T_j^{-1}(T-1)^{-1} h^{-d} \left[\varphi(\mathbf{v})-\hat{\varphi}(\mathbf{v})\right] \sum_{t=2}^{T}\sum_{s=2}^{T} \mathbf{e}_1' \tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1} \Xi \\
&\quad \times \left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) Z_s(\mathbf{u}) - T_j^{-1}(T-1)^{-1} h^{-d} \left[\varphi(\mathbf{v})-\hat{\varphi}(\mathbf{v})\right] \\
&\quad \times \sum_{t=2}^{j}\sum_{s=2}^{T} \mathbf{e}_1' \tilde{\mathbf{S}}\left(\mathbf{X}_{t-1}\right)^{-1} \Xi \left(\frac{\mathbf{X}_{s-1}-\mathbf{X}_{t-1}}{h}\right) Z_s(\mathbf{u})\, \psi_{t-j}(\mathbf{v}) \Bigg\} \left[1+o_P(1)\right] \\
&= \left[\hat{B}_{221j}(\mathbf{u},\mathbf{v}) + \hat{B}_{222j}(\mathbf{u},\mathbf{v})\right]\left[1+o_P(1)\right], \quad \text{say.} \qquad \textbf{(A.30)}
\end{aligned}
$$

To show Lemma A.2, it suffices to show that

$$
p^{-1/2} \sum_{j=1}^{T-1} k^2(j/p)\, T_j \iint \left|\hat{B}_{2abj}(\mathbf{u},\mathbf{v})\right|^2 dW(\mathbf{u})\, dW(\mathbf{v}) = o_P(1), \quad \text{for } a,b = 1,2. \qquad \textbf{(A.31)}
$$

The proof of (A.31) is similar to that of (A.12), (A.17), (A.25), and (A.26) in Lemma A.1, with the fact that $\mathrm{E}\left|\varphi(\mathbf{v})-\hat{\varphi}(\mathbf{v})\right|^4 \le C T_j^{-2}$ given Assumption 1. ■

Proof of Proposition A.2. Given the decomposition in (A.3), we have

$$\left|[\hat{\Gamma}_j(\mathbf{u},\mathbf{v})-\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})]\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})^*\right| \le \sum_{a=1}^{2} |\hat{B}_{aj}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|, \tag{A.32}$$

where the $\hat{B}_{aj}(\mathbf{u},\mathbf{v})$ are defined in (A.3).

For $a=1$, by (A.5) and the triangular inequality, we have

$$\begin{aligned}
&p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint \left|\hat{B}_{1j}(\mathbf{u},\mathbf{v})\right|\left|\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})\right| dW(\mathbf{u})dW(\mathbf{v})\\
&\quad\le p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\hat{B}_{111j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})\\
&\qquad+p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\hat{B}_{112j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})\\
&\qquad+.p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\hat{B}_{121j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})\\
&\qquad+p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\hat{B}_{122j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v}).
\end{aligned} \tag{A.33}$$

For the first term in (A.33), we have

$$\begin{aligned}
&p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint \left|\hat{B}_{111j}(\mathbf{u},\mathbf{v})\right|\left|\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})\right| dW(\mathbf{u})dW(\mathbf{v})\\
&\quad\le p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\hat{B}_{1111j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})\\
&\qquad+p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\hat{B}_{1112j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})\\
&\quad= O_P\left(p^{-1/2}T^{1/2}h^{r+1}\right)+O_P\left(p^{1/2}h^{r+1}\right)=o_P(1),
\end{aligned} \tag{A.34}$$

where we have used Assumptions 1 and 4, (A.9), and the fact that $\mathrm{E}|\tilde{\Gamma}_j(u,v)|^2 \le CT_j^{-1}$ under $\mathbb{H}_0$.

For the second term in (A.33), we have

$$\begin{aligned}
&p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint \left|\hat{B}_{112j}(\mathbf{u},\mathbf{v})\right|\left|\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})\right| dW(\mathbf{u})dW(\mathbf{v})\\
&\quad\le p^{-1/2}\sum_{j=1}^{T-1} k^2(j/p)T_j \iint |\hat{B}_{1121j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})
\end{aligned}$$

$$+p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint|\hat{B}_{1122j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})$$

$$=O_P\left(p^{-1/2}T^{1/2}h^{r+1}\right)+O_P\left(p^{1/2}h^{r+1}\right)=o_P(1), \tag{A.35}$$

where we have used Assumptions 1 and 4 and (A.9).

For the third term in (A.33), we have

$$p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{12j}(\mathbf{u},\mathbf{v})\right|\left|\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})\right|dW(\mathbf{u})dW(\mathbf{v})$$

$$\leq p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint|\hat{B}_{121j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})$$

$$+p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint|\hat{B}_{122j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})$$

$$=O_P\left(p^{-1/2}T^{(1-\nu)/2\nu}h^{-d/\nu}\right)+O_P\left(p^{-1/2}T^{-1}h^{-d/2}\right)=o_P(1), \tag{A.36}$$

where we have used Assumptions 1 and 4 and (A.9).

For the fourth term in (A.33), we have

$$p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint\left|\hat{B}_{114j}(\mathbf{u},\mathbf{v})\right|\left|\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})\right|dW(\mathbf{u})dW(\mathbf{v})$$

$$\leq p^{-1/2}\sum_{j=1}^{T-1}k^2(j/p)T_j\iint|\hat{B}_{1141j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})$$

$$p^{-1/2}+\sum_{j=1}^{T-1}k^2(j/p)T_j\iint|\hat{B}_{1142j}(\mathbf{u},\mathbf{v})||\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|dW(\mathbf{u})dW(\mathbf{v})$$

$$=O_P\left(p^{-1/2}h^{-d/\nu}\right)+O_P\left(p^{1/2}T^{-1}h^{-3d/\nu}\right)$$

$$+O_P\left(p^{3/2}T^{-1/2}\right)+O_P\left(pT^{-1}h^{-d/2}\right)=o_P(1), \tag{A.37}$$

where we have used Assumptions 1 and 4, (A.9), and the fact that $\mathrm{E}|\tilde{\Gamma}_j(u,v)|^2\leq CT_j^{-1}$ given Assumption 1.

For $a=2$, similar arguments apply. ■

Proof of Theorem A.2. The proof is similar to that of Theorem A.2 of Chen and Hong (2009). ■

Proof of Theorem A.3. The proof is similar to that of Theorem A.3 of Chen and Hong (2009). ■

Proof of Theorem 2. The proof of Theorem 2 consists of the proofs of Theorems A.4 and A.5 below. ■

THEOREM A.4. *Under the conditions of Theorem 2,* $(p^{1/2}/T)(\hat{M}-\tilde{M}) \xrightarrow{p} 0.$

THEOREM A.5. *Under the conditions of Theorem 2,*

$$(p^{1/2}/T)\tilde{M} \xrightarrow{p} D^{-1/2} \iint \int_{-\pi}^{\pi} |F(\omega,\mathbf{u},\mathbf{v}) - F_0(\omega,\mathbf{u},\mathbf{v})|^2 d\omega dW(\mathbf{u}) dW(\mathbf{v}).$$

Proof of Theorem A.4. It suffices to show that

$$T^{-1} \iint \sum_{j=1}^{T-1} k^2(j/p) T_j \left[|\hat{\Gamma}_j(\mathbf{u},\mathbf{v})|^2 - |\tilde{\Gamma}_j(\mathbf{u},\mathbf{v})|^2 \right] dW(\mathbf{u}) dW(\mathbf{v}) \xrightarrow{p} 0, \qquad \textbf{(A.38)}$$

$p^{-1}(\hat{C}-\tilde{C}) = O_P(1)$, and $p^{-1}(\hat{D}-\tilde{D}) \xrightarrow{p} 0$, where $\tilde{C}$ and $\tilde{D}$ are defined in the same way as $\hat{C}$ and $\hat{D}$ in (2.19), with $\hat{\varphi}\left(\mathbf{u}|\mathbf{X}_{t-1}\right)$ replaced by $\varphi\left(\mathbf{u}|\mathbf{X}_{t-1}\right)$. Since the proofs for $p^{-1}(\hat{C}-\tilde{C}) = O_P(1)$ and $p^{-1}(\hat{D}-\tilde{D}) \xrightarrow{p} 0$ are straightforward, we focus on the proof of (A.38). From (A.9), the Cauchy-Schwarz inequality, and the fact that $T^{-1} \iint \sum_{j=1}^{T-1} k^2(j/p) T_j |\tilde{\Gamma}_j(u,v)|^2 dW(u) dW(\mathbf{v}) = O_P(1)$ as is implied by Theorem A.5 (the proof of Theorem A.5 does not depend on Theorem A.4), it suffices to show that $T^{-1}\hat{A}_1 \xrightarrow{p} 0$, where $\hat{A}_1$ is defined as in (A.2). This is very similar to the proof of Proposition A.1, and hence it completes the proof for Theorem A.4. ■

Proof of Theorem A.5. The proof is similar to that of Theorem A.5 of Chen and Hong (2009). ■