Testing for the Markov Property in Time Series

Bin Chen
Department of Economics
University of Pittsburgh
and
Yongmiao Hong
Department of Economics and Statistical Science
Cornell University,
Wang Yanan Institute for Studies in Economics
Xiamen University

December, 2007

We would like to thank Oliver Linton, James MacKinnon, Katsumi Shimotsu, George Tauchen and seminar participants at Queen's University for their comments and discussions. Any remaining errors are solely ours. Yongmiao Hong thanks the support from Cheung Kong Scholarship from Chinese Ministry of Education and Xiamen University. Correspondences: Bin Chen, Department of Economics, University of Pittsburgh, Pittsburgh, PA 15213; Email: binchen@pitt.edu; Yongmiao Hong, Department of Economics & Department of Statistical Science, Cornell University, Ithaca, NY 14850, USA, and Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen 361005, China; Email: yh20@cornell.edu.

Testing for the Markov Property in Time Series

Abstract

The Markov property is a fundamental property in time series analysis and is often assumed in economic and financial modelling. We develop a 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 exist) 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 non-Markov alternatives in finite samples. We apply the test to a number of high-frequency financial time series and find strong evidence against the Markov property.

Key words: Markov property, Conditional characteristic function, Generalized cross-spectrum, Smoothed nonparametric bootstrap

JEL No: C1 C4 G0

1. INTRODUCTION

The Markov property is a fundamental property in time series analysis and is often a maintained assumption in economic and financial modelling. Testing for the validity of the Markov property has important implications in economics, finance as well as time series analysis. In economics, Markov decision processes (MDP), which are based on the Markov assumption, provide a broad framework for modelling 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, Hansen and Singleton 1983, Mehra and Prescott 1985), economic growth (Uzawa 1965, Romer 1986, 1990, Lucas 1988), optimal taxation (Lucas and Stokey 1983, Zhu 1992), and equilibrium business cycles (Kydland and Prescott 1982). In the MDP framework, an optimal decision rule can be found within the subclass of non-randomized 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 (e.g., Mizutani and Dreyfus 2004, Aviv and Pazgal 2006). 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 among most continuous-time models. It is well known that stochastic integrals yield Markov processes. In modelling 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 the interest rate dynamics. 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, an 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 must 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. As a result, 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 prices series alone, which itself is non-Markov, 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, Amaro de Matos and Rosario (2000) and Platen and Rebolledo (1996) propose equilibrium models, which assume that market makers can take advantage of their superior information on trade orders and set different bid and ask 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 reflects reality more appropriately.

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.

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 Skϕderberg (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 which are not Markov but whose first order transition probabilities nevertheless satisfy the Chapman-Kolmogorov equation. Ait-Sahalia's (1997) test has no power against these non-Markov processes.

Amaro de Matos and Fernandes (2007) extend the smoothed nonparametric density approach proposed by Fernandes and Flôres (2004) to test whether discretely recorded observations of a continuous-time process are consistent with the Markov property. 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 an omnibus test for the Markov property. The characteristic function has been widely used in time series analysis and financial econometrics (e.g., Feuerverger and McDunnough 1981, Epps 1987, 1988, Feuerverger 1990, Hong 1999, Singleton 2001, Jiang and Knight 2002, Knight and Yu 2002, Chacko and Viceira 2003, Carrasco, Chernov, Florens and Ghysels 2007, and Su and White 2007a). 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 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 "curse of dimensionality" problem. This improves upon the existing tests, which can only check

[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 (2007a, 2007b) check conditional independence by the Hellinger distance and empirical characteristic function respectively. These tests can be used to test the Markov property. However, they will encounter the "curse of dimensionality" problem because the Markov property implies that conditional independence must hold for infinite number of lags.

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.

Fourth, unlike tests based on characteristic functions in the statistical literature (e.g., Epps and Pulley 1983), which often have nonstandard asymptotic distributions, our test statistic has a convenient null asymptotic $N(0,1)$ distribution.

In Section 2, we describe the hypotheses of interest and propose a novel approach to testing for the Markov property. We derive the asymptotic distribution of the proposed test statistic in Section 3, and discuss its asymptotic power in Section 4. In Section 5, we use Paparoditis and Politis' (2000) smoothed nonparametric transition-based bootstrap procedure to obtain the critical values of the test in finite samples and examine the finite sample performance of the test. In Section 6, we apply our test to stock prices, interest rates and foreign exchange rates and document strong evidence against the Markov property with all three financial time series. Section 7 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 will use C to denote a generic bounded constant, $\|\cdot\|$ for the Euclidean norm, and A^* for the complex conjugate of A.

2. HYPOTHESES OF INTEREST 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 follows:

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

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

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

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

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

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

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 has no power against 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

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

which is implied by $\mathbb{H}_0$. By choosing $j=1$, Amaro de Matos and Fernandes (2007) check

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

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 the existing tests essentially check the conditional independence of

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

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 $\mathcal{I}_t$. There will be inevitably 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,

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

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

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

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 follows:

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

versus the alternative hypothesis

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

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

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

Then the Markov property is equivalent to the following MDS characterization

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

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

$$\exp(i\mathbf{u}'\mathbf{X}_{t+1}) = E[\exp(i\mathbf{u}'\mathbf{X}_{t+1})|\mathbf{X}_t] + Z_{t+1}(\mathbf{u}) = \varphi(\mathbf{u}|\mathbf{X}_t) + Z_{t+1}(\mathbf{u}).$$

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

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

for for all $\mathbf{u}$ near 0.[2] Thus, checking (2.6) is equivalent to checking whether all conditional moments of $\mathbf{X}_{t+1}$ (if exist) are Markov. Nevertheless, the use of (2.6) itself does not require any moment conditions on $\mathbf{X}_{t+1}$.

It is not a trivial task to check (2.6). First, the MDS property in (2.6) must hold for all $\mathbf{u} \in \mathbb{R}^d$,

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

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 and 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.[3] 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

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

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

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

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

[3] This is not a trivial extension since we use nonparametric estimation in the first stage.

cross-spectrum $F(\omega, \mathbf{u}, \mathbf{v})$ becomes a "flat" spectrum as a function of frequency ω:

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

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.

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

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

where the estimated generalized residual

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

$\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$ 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 estimation methods are available. We use the most popular kernel method, mainly due to its simplicity and intuitive appeal.

To ensure that the nonparametric CCF estimator $\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$ has a fast convergence rate so that the estimation of $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$ has no impact on the generalized cross-spectral density estimation asymptotically, we use a higher order kernel to estimate $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$. For this, we introduce a rth order kernel K, such that $\int_{-\infty}^{\infty} K(u)\,du = 1$, $\int_{-\infty}^{\infty} u^a K(u)\,du = 0$, $\int_{-\infty}^{\infty} u^r K(u)\,du = B_K < \infty$ and $B_K \neq 0$. Examples of higher order kernels include: $K_4(u) = \frac{1}{2}(3 - u^2)\,\phi(u)$, $K_6(u) = \frac{1}{8}(15 - 10u^2 + u^4)\,\phi(u)$, $K_8(u) = \frac{1}{48}(105 - 105u^2 + 21u^4 - u^6)\phi(u)$, where $\phi(u) = (2\pi)^{-1/2}\exp(-\frac{1}{2}u^2)$. Then the Nadaraya-Watson estimator for $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$ is

$$\hat{\varphi}(\mathbf{u}|\mathbf{x}) = \frac{\sum_{s=2}^T e^{i\mathbf{u}'\mathbf{X}_s}\mathbf{K}_h(\mathbf{x} - \mathbf{X}_{s-1})}{\sum_{s=2}^T \mathbf{K}_h(\mathbf{x} - \mathbf{X}_{s-1})}, \tag{2.12}$$

where $\mathbf{K}_h(\mathbf{x} - \mathbf{X}) = \prod_{a=1}^d K_h(x_a - X_{at}) = \prod_{a=1}^d \left[h^{-1}K\left(\frac{x_a - X_{at}}{h}\right)\right]$ and $h = h(T)$ is the bandwidth. The regression estimator $\hat{\varphi}(\mathbf{u}|\mathbf{x})$ only involves a d-dimensional smoothing, thus enjoying some advantages over the existing nonparametric density approaches which involve a $2d$ or $3d$ dimensional smoothing.

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

$$\hat{F}_0(\omega, \mathbf{u}, \mathbf{v}) = \frac{1}{2\pi}\hat{\Gamma}_0(\mathbf{u}, \mathbf{v}), \qquad \omega \in [-\pi, \pi], \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^d. \tag{2.13}$$

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

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

where $p = p(T) \to \infty$ is a bandwidth or 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 $K(\cdot)$ in (2.12). Most commonly used kernels discount higher order lags. Examples of commonly used $k(z)$ include the Bartlett kernel

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

the Parzen kernel

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

and the Quadratic-Spectral kernel

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

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

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

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.[4] An example of $W(\cdot)$ is the multivariate independent $N(\mathbf{0},\mathbf{I})$ 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 proposed omnibus test statistic for $\mathbb{H}_0$ against $\mathbb{H}_A$ is an appropriately standardized version of (2.18), namely,

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

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

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

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 ECF of $\{\mathbf{X}_t\}$. The factors $\hat{C}$ and $\hat{D}$ are approximately the mean and variance of the quadratic form in (2.18). 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 d of $\mathbf{X}_t$ is large. Alternatively, one can only use a finitely many 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

[4] If $W(\mathbf{u})$ is differentiable, then its derivative $(\partial/\partial u_a)W(\mathbf{u})$ is an even function of u_a for $a=1,2,...d$.

dramatically reduce the computational cost but it may lead to some power loss. We will examine this issue by simulation studies.

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 involves $2d$- 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 statistic $\hat{M}$, we impose the following regularity conditions.

Assumption A.1: *(i)* $\{\mathbf{X}_t\}$ *is a strictly stationary* β*-mixing process with mixing coefficient* $\beta(j) = O(j^{-\nu})$ *for some constant* $\nu > 3$; *(ii) the marginal density* $g(\mathbf{x})$ of $\mathbf{X}_t$ *is bounded away from* $\mathbf{0}$ *and has at least rth order partial derivatives for some integer* $r > 0$.

Assumption A.2: *For each sufficiently large integer q, there exists a* $q-$*dependent stationary process* $\{\mathbf{X}_{qt}\}$, *such that* $E\|\mathbf{X}_t - \mathbf{X}_{qt}\|^2 \leq Cq^{-\delta}$ *for some constant* $\delta \geq 1$ *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 A.3: $K(\cdot)$ *is a rth order, bias-reduction kernel, satisfying* $\int_{-\infty}^{\infty} K(u)\,du = 1$, $\int_{-\infty}^{\infty} u^a K(u)\,du = 0$, for $0 < a \leq r-1$, and $\int_{-\infty}^{\infty} u^r K(u)\,du < \infty$, *where r is the same as in Assumption A.1.*

Assumption A.4: *(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* $b > \frac{1}{2}$ *as* $z \to \infty$.

Assumption A.5: $W : \mathbb{R}^d \to \mathbb{R}^+$ *is a nondecreasing weighting function that weighs sets symmetric about the origin equally, with* $\int dW(\mathbf{u}) < \infty$ *and* $\int \|\mathbf{u}\|^4 dW(\mathbf{u}) < \infty$.

Assumption A.1 and A.2 are regularity conditions on the DGP of $\{\mathbf{X}_t\}$. Assumption A.1(i) restricts the degree of temporal dependence of $\{\mathbf{X}_t\}$. We say that $\{\mathbf{X}_t\}$ is β-mixing (absolutely regular) if

$$\beta(j) = \sup_{s\geq 1} E\left[\sup_{A\in\mathcal{F}_{s+j}^{\infty}} |P(A|\mathcal{F}_1^s) - P(A)|\right] \to 0,$$

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 A.1(i) holds for many well-known processes such as linear stationary ARMA processes and a large class of processes implied by numerous nonlinear models, including bilinear, nonlinear AR, and ARCH-type models (Fan and Li, 1999). Ait-Sahalia, Fan and Peng (2006), Amaro de Matos and Fernandes (2007) and Su and White (2007a, 2007b) also impose β-mixing conditions. Our mixing condition is weaker than Amaro de Matos and Fernandes' (2007) and Su and White's (2007b). They assume a β-mixing condition with a geometric decay rate.

Assumption A.1(ii) first appears restrictive as it rules out some most commonly used probability densities, such as $N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. This allows us to focus on the essentials and still maintain a relatively straightforward treatment. To satisfy such an assumption, it is a common practice to exclude data in the tails. This, however, leads to information loss. Tails may be particularly informative and interesting for financial time series.

It is well-known that the Markov property is invariant to any strictly monotonic transformation.[5] One can exploit such a property to ensure that $g(\cdot)$ is bounded away from zero from below. Let $\mathbf{Y}_t$ have the cumulative density function (CDF) $\tilde{G}(\cdot)$ with density $\tilde{g}(\cdot)$ and let $L(\cdot)$ be a prespecified CDF with density $l(\cdot)$. Then $\mathbf{X}_t \equiv L(\mathbf{Y}_t)$ has support on $\mathbb{R}^d$ and the CDF of $\mathbf{X}_t$ is given by $G(\mathbf{x}) = \tilde{G}[L^{-1}(\mathbf{x})]$, $\mathbf{x} \in \mathbb{R}^d$. It follows that

$$g(\mathbf{x}) \equiv \frac{\partial^d G(\mathbf{x})}{\partial x_1 ... \partial x_d} = \tilde{g}\left[L^{-1}(\mathbf{x})\right] \left|\det J\left[L^{-1}(\mathbf{x})\right]\right|^{-1},$$

where $J(\cdot)$ is the Jacobian matrix. To ensure $\min_{\mathbf{x}\in\mathbb{R}^d} g(\mathbf{x}) \geq c > 0$, it suffices that $|\det J(\mathbf{y})| \leq c^{-1}\tilde{g}(\mathbf{y})$, $\mathbf{y} \in \mathbb{R}^d$.[6] No information would be lost in the transformation. In fact, the condition that $g(\mathbf{x}) \geq c > 0$ is made for simplicity of the asymptotic analysis. It seems plausible that one could allow $g(\mathbf{x}) \to 0$ at the end points with a sufficiently slow rate, and our theory would continue to hold under strengthened conditions on the bandwidth h used in kernel regression estimation of CCF. As the involved technicality would be quite complicated and would detract from our main goal, we do not pursue this here. However, we will use simulation to examine the consequence of allowing $g(\mathbf{x}) \to 0$ at the end points.

Assumption A.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. 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 A.2 holds trivially. Assumption A.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 $\{\mathbf{X}_t\}$:

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

[5] Rosenblatt (1971, Ch. III) provides conditions under which functions of a Markov process are Markov.
[6] When $d = 1$, it boils down to $l(\mathbf{y}) \leq c^{-1}\tilde{g}(\mathbf{y})$.

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

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

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

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

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

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

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

Thus Assumption A.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

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

Thus Assumption A.2 holds.

Assumption A.3 imposes regularity conditions on the kernel function $k(\cdot)$ used for nonparametric regression estimation of CCF. Bierens (1987), Pagan and Ullah (1999) and Li and Racine (2007) provide a discussion of how to construct specific kernels satisfying these conditions. The class of higher order kernels allows for reducing the bias of the marginal density estimator $\hat{p}(\mathbf{x})$ and therefore obtains a faster rate of convergence. The smoothness of $p(\mathbf{x})$ as measured by the derivative order r determines how much the bias can be reduced with. These bias-reduction kernels have been widely used in the literature (e.g., Robinson 1988, 1989, Andrews 1995).

Assumption A.4 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 A.4 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 A.5 imposes mild conditions on the prespecified weighting function $W(\cdot)$. Any CDF with finite fourth moments satisfies Assumption A.6. Note that $W(\cdot)$ need not be continuous. This provides a convenient way to implement our tests, because we can avoid relatively high dimensional numerical integrations by using finitely many numbers of grid points for $\mathbf{u}$ and $\mathbf{v}$.

Theorem 1: *Suppose Assumptions A.1−A.5 hold, and* $p = cT^{\lambda}$ *for* $\frac{2d}{2r\nu+d} < \lambda < (3 + \frac{1}{4b-2})^{-1}$ *and* $0 < c < \infty$, $h = cT^{-\delta}$, $\delta \in \left(\frac{2-\lambda}{4r}, \min(\frac{\lambda\nu}{2d}, \frac{2-\lambda}{2d})\right)$. *Then under* $\mathbb{H}_0$,

$$\hat{M} \to^d N(0,1) \text{ as } T \to \infty.$$

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 $\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$ by $\varphi(\mathbf{u}|\mathbf{X}_{t-1})$ has no impact on the asymptotic distribution of $\hat{M}$ under H_0. However, $\hat{\varphi}(\mathbf{u}|\mathbf{X}_{t-1})$ may have substantial impact on the finite sample size performance of the $\hat{M}$ test. To overcome such adverse impact, we will use Paparoditis and Politis' (2000) 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.

We note that our condition on bandwidth h allows the optimal bandwidth rates for estimating CCF. Thus, data-driven choices of bandwidth, which usually balance the variance and squared bias, can be used. In contrast, Ait-Sahali (1997) and Su and White (2007a) require an undersmoothing procedure to ensure that the squared bias vanishes to zero faster than the variance. Such undersmoothing rules out asymptotically optimal data-driven choices of bandwidth h.

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 behavior of $\hat{M}$ under $\mathbb{H}_A$ in (2.2).

Theorem 2: *Suppose Assumption A.1 and A.3- A.5 hold, and* $p = cT^{\lambda}$ *for* $\frac{2d}{2r\nu+d} < \lambda < (3+\frac{1}{4b-2})^{-1}$ *and* $0 < c < \infty$, $h = cT^{-\delta}$, $\delta \in \left(\frac{2-\lambda}{4r}, \min(\frac{\lambda\nu}{2d}, \frac{1}{d})\right)$. *Then under* $\mathbb{H}_A$*, and as* $T \to \infty$,

$$\begin{aligned} \frac{p^{\frac{1}{2}}}{T}\hat{M} &\rightarrow {}^{p} \frac{1}{\sqrt{D}} \sum_{j=1}^{\infty} \int\int \left|\Gamma_j\left(\mathbf{u},\mathbf{v}\right)\right|^2 dW\left(\mathbf{u}\right) dW\left(\mathbf{v}\right) \\ &= \frac{1}{\sqrt{D}} \int\int\int_{-\pi}^{\pi} \left|F\left(\omega,\mathbf{u},\mathbf{v}\right) - F_0\left(\omega,\mathbf{u},\mathbf{v}\right)\right|^2 d\omega dW\left(\mathbf{u}\right) dW\left(\mathbf{v}\right), \end{aligned}$$

where

$$\begin{aligned} D &= 4\pi \int_0^{\infty} k^4\left(z\right) dz \int\int \left|\Sigma_0\left(\mathbf{u}_1,\mathbf{u}_2\right)\right|^2 dW\left(\mathbf{u}_1\right) dW\left(\mathbf{u}_2\right) \\ &\quad \times \int\int\int_{-\pi}^{\pi} \left|L\left(\omega,\mathbf{v}_1,\mathbf{v}_2\right)\right|^2 d\omega dW\left(\mathbf{v}_1\right) dW\left(\mathbf{v}_2\right), \end{aligned}$$

and $L\left(\omega,\mathbf{u},\mathbf{v}\right) = (2\pi)^{-1}\sum_{j=-\infty}^{\infty} \Omega_j\left(\mathbf{u},\mathbf{v}\right) e^{-ij\omega}$, $\Omega_j\left(\mathbf{u},\mathbf{v}\right) = cov\left(e^{i\mathbf{u}'\mathbf{X}_t}, e^{i\mathbf{v}'\mathbf{X}_{t-|j|}}\right)$ *and* $\Sigma_0\left(\mathbf{u},\mathbf{v}\right) = cov\left[Z_t\left(\mathbf{u}\right), Z_t\left(\mathbf{v}\right)\right]$.

The function $L\left(\omega,\mathbf{u},\mathbf{v}\right)$ 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 $L\left(\omega,\mathbf{u},\mathbf{v}\right)$ is due to the fact that the conditioning variable $\{\exp\left(i\mathbf{v}'\mathbf{X}_{t-|j|}\right)\}$ 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 $E\left[Z_t(\mathbf{u})|\mathbf{X}_{t-j}\right] = 0$ *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 $E\left[Z_t(\mathbf{u})|\mathbf{X}_{t-j}\right] \neq 0$ at some lag $j > 0$. Then we have $\int\int \left|\Gamma_j\left(\mathbf{u},\mathbf{v}\right)\right|^2 dW\left(\mathbf{u}\right) dW\left(\mathbf{v}\right) > 0$ for any weighting function $W\left(\cdot\right)$ that is positive, monotonically increasing and continuous, with unbounded support on $\mathbb{R}^d$. As a consequence, $P[\hat{M} > C\left(T\right)] \to 1$ 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 $E\left[Z_t(\mathbf{u})|\mathbf{X}_{t-j}\right] \neq 0$ 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 the 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 avoid 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 $\mathbf{v}$.

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 consider a sequence of tests based on the derivatives of the nonparametric regression residual $Z_t(\mathbf{u})$ at the origin $\mathbf{0}$:

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

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, the choices of $\mathbf{m} = 1,2,3,4$ corresponds to tests for departures of the Markov property in the first fourth conditional moments respectively. For each $\mathbf{m}$, the resulting test statistic is given by:

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

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

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

the centering and scaling factors

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

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

and

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

These derivative tests may provide additional useful information on the possible sources of the violation of the Markov property. Moreover, 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 $E\left[u'\left(C_{t+1}\right)|\mathcal{I}_t\right] = u'\left(C_t\right)$, where $u'(C_t)$ is the marginal utility of consumption C_t. This can be viewed as the Markov property in mean. The derivative test $\hat{M}\left(1\right)$ can be used to test this implication.

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.

5. FINITE SAMPLE PERFORMANCE

Theorem 1 provides the asymptotic null distribution of $\hat{M}$. Consequently, 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, its size in finite samples may differ significantly from the asymptotic significance level. Our analysis suggests that the asymptotic theory may not work well even for relatively large samples, because the asymptotically negligible higher order terms in $\hat{M}$ are close in order of magnitude to the dominant U-statistic, which 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. Our simulation study shows that $\hat{M}$ displays severer underrejection under $\mathbb{H}_0$. On the other hand, we examine the finite sample performance of an infeasible $\hat{M}$ test by replacing the estimated generalized residual $\hat{Z}_t(\mathbf{u})$ by the true generalized residual. 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), Fan *et al.* (2006) and Hong and White (2005) in other contexts.

To overcome this problem, we use Paparoditis and Politis' (2000) smoothed nonparametric conditional density bootstrap procedure to more accurately approximate the finite-sample null distribution of $\hat{M}$.[7] 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

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

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

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

  where $K(\cdot)$ and h are the same as those used in $\hat{M}$;[8]

- 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 $K(\cdot)$ 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 $p_b \equiv B^{-1}\sum_{l=1}^{B} \mathbf{1}(\hat{M}_l^b > \hat{M})$. To obtain accurate bootstrap p-values, B must be sufficiently large.

We emphasize that the same kernel $K(\cdot)$, the same bandwidth h should be used in $\hat{g}(\mathbf{x}|\mathbf{X}_{t-1})$, $\hat{M}$ and $\hat{M}^b$. This helps obtain a better size for our test in finite samples. Because the nonparametric transition density estimator $\hat{g}(\mathbf{x}|\mathbf{X}_{t-1})$ in (5.1), is consistent for the true transition density of the process $\{\mathbf{X}_t\}$ under $\mathbb{H}_0$, the bootstrap distribution of a test statistic based on the bootstrap sample will mimic the distribution of the test based on the original sample (Paparoditis and Politis 2000, and Horowitz 2003). Smoothed nonparametric bootstraps have been used to improve finite sample performance in hypothesis testing. For example, Su and White (2007a, 2007b) apply Paparoditis and Politis' (2000) procedure in testing for conditional independence. Amaro de Matos and Fernandes (2007) use Horowitz's (2003) Markov conditional bootstrap procedure in testing for the Markov property. Paparoditis and Politis' (2000) procedure is very similar to Horowitz's (2003), except that Paparoditis and Politis (2000) generate bootstrap samples from $\hat{g}(\mathbf{x}|\mathbf{X}_{t-1})$ and Horowitz (2003) generates bootstrap samples from $\hat{g}(\mathbf{x}|\mathbf{X}_{t-1}^b)$. These two procedures can be applied to our test. We expect that the latter is more computationally expensive since bootstrap observations have to be generated sequentially.

[7] For the application of the bootstrap in econometrics, see (e.g.) Horowitz (2001).

[8] Bootstrap samples can be generated by applying the inverse-distribution method to a fine grid of points.

To examine the sizes of our test under $\mathbb{H}_0$, we consider two Markov DGPs:

$$\text{DGP S.1 [AR(1)]:} \qquad X_t = 0.5X_{t-1} + \varepsilon_t,$$

$$\text{DGP S.2 [ARCH(1)]:} \quad \begin{cases} X_t = h_t^{\frac{1}{2}}\varepsilon_t \\ h_t = 0.1 + 0.1X_{t-1}^2, \end{cases}$$

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

To examine the power of our test with the smoothed bootstrap procedure, we consider the following non-Markovian DGPs:

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

$$\text{DGP P.2 [GARCH(1,1)]:} \qquad \begin{cases} X_t = h_t^{\frac{1}{2}}\varepsilon_t \\ h_t = 0.1 + 0.1X_{t-1}^2 + 0.8h_{t-1}, \end{cases}$$

$$\text{DGP P.3 [Markov Regime-Switching]:} \qquad X_t = \begin{cases} 0.7X_{t-1} + \varepsilon_t, & \text{if } S_t = 0, \\ -0.3X_{t-1} + \varepsilon_t, & \text{if } S_t = 1, \end{cases}$$

$$\text{DGP P.4 [Markov Regime-Switching ARCH]:} \qquad \begin{cases} X_t = \begin{cases} \sqrt{h_t}\varepsilon_t, & \text{if } S_t = 0, \\ 3\sqrt{h_t}\varepsilon_t, & \text{if } S_t = 1, \end{cases} \\ h_t = 0.1 + 0.3X_{t-1}^2, \end{cases}$$

where $\varepsilon_t \sim i.i.d.N(0,1)$, and in DGP P.3 and P.4, S_t is a latent state variable that follows a two-state Markov chain with transition probabilities $P(S_t = 1|S_{t-1} = 0) = P(S_t = 0|S_{t-1} = 1) = 0.9$. DGP P.3 and P.4 are Markov Regime-Switching model and Markov 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 the latent state variable S_t changes the Markov property of AR and ARCH processes. The knowledge of $\mathbf{X}_{t-1}$ is not sufficient to summarize all relevant informations in $\mathcal{I}_{t-1}$ that is useful to predict the future behavior of the time series process. The departure from the Markov property comes from the conditional mean in DGP P.1 and P.3 and from the conditional variance in DGP P.2 and P.4.

Throughout, we consider two sample sizes: $T = 100, 250$. 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 in the simulation study, we generate $\mathbf{u}$ and $\mathbf{v}$ from a $N(0,1)$ distribution, with each $\mathbf{u}$ and $\mathbf{v}$ having 30 symmetric grid points in $\mathbb{R}$ respectively.[9] We use the Bartlett kernel in (2.14), which has bounded support and is computationally efficient. Our simulation experience suggests that the

[9] 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.

choices of $W(\cdot)$ and $k(\cdot)$ have little impact on both the size and power of the tests.[10] 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 the preliminary bandwidth $\bar{p}$ on the size and power of the tests, we consider $\bar{p}$ in the range of 5 to 20. Following Robinson (1991) and Su and White (2007a), we use a fourth order kernel $K(u) = \frac{1}{2}(3 - u^2)\phi(u)$, where $\phi(u)$ is the N(0,1) density. For simplicity, we choose $h = \hat{S}_X T^{-\frac{1}{5}}$, where $\hat{S}_X$ is the sample standard deviation of $\{X_t\}_{t=1}^T$.

Tables 1 reports the bootstrap sizes and powers of $\hat{M}$ at the 10% and 5% nominal significance levels under DGPs S.1-2 and P.1-4. The test has reasonable sizes under both DGPs S.1 and S.2 for sample size $T = 100$, at both the 10% and 5% levels. Under DGP S.1 (AR(1)), $\hat{M}$ tends to overreject a little but the overrejection is not excessive. Under DGP S.2 (ARCH(1)), the empirical levels are very close to the nominal levels, especially at the 5% level. 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.[11]

Under DGPs P.1-4, $\mathbf{X}_t$ is not Markov and our test has reasonable power to detect them. The rejection rates are around 20% at the 5% level when $T = 100$ and reach 40% when $T = 250$ under DGPs P.1, 3 and 4 (MA(1), Markov chain regime-switching and Markov chain regime-switching ARCH(1)). Under DGP P.2 (GARCH(1,1)), the rejection rate is a bit smaller but still reaches 15% at the 5% level when $T = 250$, and increases with the sample size T.

In summary, the test with the smoothed bootstrap procedure delivers reasonable size and omnibus power against various non-Markov alternatives in small samples.

6. DO FINANCIAL TIME SERIES FOLLOW MARKOV PROCESSES?

As documented by Hong and Li (2005), such popular spot interest rates continuous-time models as Vasicek (1977), Cox, Ingersoll and Ross (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 the spot interest rate processes. Although some works are still going on to add the richness of model specification in terms of jumps and functional forms, the models remain 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 (2005). If this is indeed the case, then 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

[10]We have tried the Parzen kernel for $k(\cdot)$, obtaining similar results (not reported here).
[11]We have tried Horowitz's (2003) Markov conditional bootstrap. The results, not reported here, are rather similar but it takes much longer computational time.

O'Hara (1987,1992)) is that the asset price sequence does not follow a Markov process. It is interesting to check whether real stock prices are consistent with such a conjecture.

We apply our tests to three important financial time series: stock prices, interest rates and foreign exchange rates. We use S&P500, 7-day Eurodollar rate and Japanese Yen, obtained from Datastream. Two sample periods are considered: January 1, 1988 to December 31, 2006 for a total of 4,427 daily observations, and January 1, 1998 to December 31, 2006 for a total of 2,263 observations. To remove the time trend and nonstationarity, we consider S&P500 value-weighted returns, 7-day Eurodollar rate changes and Japanese Yen returns. Figures 1-6 provides the time series plots and Table 2 reports the test statistics and bootstrap p-values for preliminary lag orders $\bar{p}$ from 10 to 20. The bootstrap $p-$values, based on $B = 500$ bootstrap iterations, are computed as described in Section 5. For both the sample periods, the test statistics are quite robust to the choice of $\bar{p}$ and have essentially zero bootstrap p-values, revealing strong evidence against the Markov property for S&P500 returns, 7-day Eurodollar rate changes and Japanese Yen returns. The full sample from 1988 yields larger statistics values than the subsample from 1998, which confirms the monotonic power of the new test.

These results cast some new thoughts on financial modelling. Although 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 modelling schemes, which allow for the non-Markov assumption, may be needed to better capture the dynamics of financial time series processes.

7. 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. Our results suggest that the Markov assumption may not be suitable for many financial time series.

REFERENCE

Aaronson, J., D. Gilat, and M. Keane (1992): "On the Structure of 1-dependent Markov Chains,".*Journal of Theoretical Probability,* 5, 545-561.

Ahn, D., R. Dittmar and A.R. Gallant (2002): "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies,* 15, 243-288.

Ahn, D. and 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.

—— (1997): "Do Interest Rates Really Follow Continuous-Time Markov Diffusions?" working paper, Princeton University.

Ait-Sahalia, Y., J. Fan and H. Peng (2006): "Nonparametric Transition-Based Tests for Diffusions," working paper, Princeton University.

Amaro de Matos, J.and M. Fernandesbes (2007): "Testing the Markov property with high frequency data," *Journal of Econometrics,* 141, 44-64.

Amaro de Matos, J. and 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. and J. Lund (1997): "Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate," *Journal of Econometrics,* 77, 343-377.

Andrews, D. W. K. (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560-596.

Aviv, Y. and 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 and 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.

—— (1987): "Kernel Estimators of Regression Functions," *Advances in Econometrics: Fifth World Congress,* Truman F.Bewley (ed.), Vol.I, New York: Cambridge University Press, 99-144.

Blume, L., D. Easley, and M. O'Hara (1994): "Market Statistics and Technical Analysis: The Role of Volume," *Journal of Finane,* 49, 153-181.

Carrasco, M., M.Chernov, J.P. Florens and E. Ghysels (2007): "Efficient Estimation of Jump Diffusions and General Dynamic Models with a Continuum of Moment Conditions," *Journal of Econometrics,* forthcoming.

Chacko, G., and 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, and A.B. Sanders (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance,* 47, 1209-1227.

Chen, B. and Y. Hong (2007): "Diagnosing Multivariate Continuous-Time Models with Application to Affine Term Structure Models," working paper, Cornell University.

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

Dai, Q., and K. Singleton (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance,* 55, 1943-1978.

Darsow, W. F., B. Nguyen, and 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.

—— (1987): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 74, 33-43.

Duffie, D. and R. Kan (1996): "A Yield-Factor Model of Interest Rates," *Mathematical Finance,* 6, 379-406.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica,* 68, 1343-1376.

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

—— (1992): "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47, 577-605.

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

Epps, T.W. and L.B. Pulley (1983): "A test for Normality Based on the Empirical Characteristic Function," *Biometrika*, 70, 723-726.

Fan, Y. and Q. Li, (1999): "Root-N-consistent Estimation of Partially Linear Time Series Models," *Journal of Nonparametric Statistics,* 11, 251-269.

Fan, Y., Q. Li and 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.

Fernandes, M. and R. G. Flôres (2004): "Nonparametric Tests for the Markov Property," Working paper, Getulio Vargas Foundation.

Feuerverger, A. and P. McDunnough (1981): "On the Efficiency of Empirical Characteristic Function Procedures," *Journal of the Royal Statistical Societh Series B,* 43, 20-27.

—— (1990): "An Efficiency Result for the ECF in Stationary Time-series Models," *Canadian Journal of Statistics*, 18, 155-161.

Gallant, A. R., D. Hsieh and G. Tauchen (1997): "Estimation of Stochastic Volatility Models with Diagnostics," *Journal of Econometrics,* 81, 159–192.

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, 1989, 357-384.
Hamilton J.D. and 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, L.P. and K. J. Singleton. (1983): "Stochastic consumption, risk aversion and the temporal behavior of asset returns," *Journal of Political Economy*, 91, 249-265.
Härdle, W. (1990): *Applied Nonparametric Regression*, Cambridge University Press: New York.
Heath, D., R. Jarrow and 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. and H. Li (2005): "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates," *Review of Financial Studies*, 18, 37-84.
Hong , Y. and H. White (2005): "Asymptotic Distribution theory for an Entropy-Based Measure of Serial Dependence," *Econometrica*, 73, 837-902.
Horowitz, J.L. (2001): "The Bootstrap in Econometrics," *Handbook of Econometrics*, 5, J. J.Heckman and E. E. Learner (ed.), 3159-3228.
—— (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.
Jarrow, R. and S. Turnbull (1995): "Pricing Derivatives on Financial Securities Subject to Credit Risk," *Journal of Finance,* 50, 53-86.
Jarrow, R., D. Lando, and S. Turnbull (1997): "A Markov Model for the Term Structure of Credit Risk Spreads," *Review of Financial Studies*, 481–523.
Jiang, G. and J. Knight (1997): "A Nonparametric Approach to the Estimation of Diffusion Processes with an Application to a Short-term Interest Rate Model," *Econometric Theory,* 13, 615-645.
Kavvathas, D. (2000):."Estimating credit rating transition probabilities for corporate bonds," working paper, Department of Economics, University of Chicago.
Kiefer, N.M. and C.E. Larson (2004): "Testing Simple Markov Structures for Credit Rating Transitions", working paper, Cornell University.
Knight, J., and J. Yu (2002): "Empirical Characteristic Function in Time Series Estimation," *Econometric Theory*, 18, 691-721.

Kydland, F.E. and E. Prescott (1982): "Time to Build and Aggregate Fluctuations," *Econometrica,* 50, 1345–70.

Lando D. and T. Skødeberg (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.

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

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

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

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

Loretan M. and 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.

—— (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3-42.

Lucas, R. and E. Prescott (1971): "Investment under Uncertainty," *Econometrica,* 39, 659-81.

Lucas, R. and N. L. Stokey (1983): "Optimal Fiscal and Monetary Policy in an Economy Without Capital," *Journal of Monetary Economics,* 12, 55-94.

Matús, F. (1996): "On Two-Block-Factor Sequences and One-Dependence," *Proceedings of the American Mathematical Society,* 124, 1237-1242.

—— (1998): "Combining M-dependence with Markovness," *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques,* 34, 407-423.

Mehra, R. and E. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics,* 15, 145–61.

Mizutani, E. and 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. and Schwert, G. W. (1990): "Testing for covariance stationarity in stock market data," *Economics Letters*, 33, 165–70.

Pagan. A. and A. Ullah (1999) *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Paparoditis, E. and D. N. Politis (2000) "The Local Bootstrap for Kernel Estimators under General Dependence Conditions," *Annals of the Institute of Statistical Mathematics*, 52, 139-159.

Platen, E and R. Rolando, (1996), "Principles for modelling financial markets," *Journal of Applied Probability* 31, 601-613.
Robinson, P. M., (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.
—— (1989): "Hypothesis Testing in Semiparametric and Nonparametric Models for Economic Time Series," *Review of Economic Studies*, 56, 511-534.
Romer, P. (1986): "Increasing Returns and Long-Run Growth," *Journal of Political Economy,* 5, 1002-1037.
—— (1990): "Endogenous Technological Change," *Journal of Political Economy*, 5, 71-102.
Rosenblatt, M. (1960): "An Aggregation Problem for Markov Chains,".*Information and Decision Processes,* Machol, R. E. (ed.), McGraw-Hill, New York, 87-92.
—— (1971) *Markov Processes. Structure and Asymptotic Behavior.* Springer-Verlag, New York-Heidelberg.
Rosenblatt, M. and D. Slepian (1962):."Nth Order Markov Chains with Every N Variables Independent," *Journal of the Society for Industrial and Applied Mathematics,* 10, 537-549.
Rust, J. (1994): "Structural Estimation of Markov Decision Processes," *Handbook of Econometrics*, 4, 3081-3143.
Sargent, T. (1987) *Dynamic Macroeconomic Theory.* Cambridge: 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. and D. Tjøstheim**,** (1993): "Nonparametric Tests of Serial Independence," in *Developments in Time Series Analysis*: The Priestley Birthday Volume, T. Subba Rao (ed.), Chapman and Hall: London, 207-229.
—— (1996): "Measures of Distance Between Densities with Application to Testing for Serial Independence," in *Time Series Analysis in Memory of E. J. Hannan*, P. Robinson and M. Rosenblatt (eds.), Springer: New York, 363-377.
Stinchcombe, M. B. and H. White (1998): "Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative," *Econometric Theory,* 14, 295-325.
Su, L. and H. White (2007a): "A Consistent Characteristic-Function-Based Test for Conditional Independence," *Journal of Econometrics,* forthcoming.
—— (2007b): "Nonparametric Hellinger Metric Test for Conditional Independence," *Econometric Theory,* forthcoming..
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.
Zhu, X. (1992): "Optimal Fiscal Policy in a Stochastic Growth Model," *Journal of Economic Theory*, 2, 250-289.

Table 1: Size and power of the test

	$T = 100$			$T = 250$		
lag	10	15	20	10	15	20
Size						
DGP S.1: AR(1)						
10%	13.2	12.6	12.2	13.8	12.2	12.4
5%	6.4	6.2	6.6	6.8	5.2	5.4
DGP S. 2: ARCH(1)						
10%	11.2	11.0	11.8	8.4	8.4	9.6
5%	7.0	7.2	6.8	4.8	5.0	5.0
Power						
DGP P.1: MA(1)						
10%	32.0	30.6	25.6	59.8	50.6	46.8
5%	20.2	18.8	17.0	46.0	39.8	36.0
DGP P.2: GARCH(1,1)						
10%	18.1	16.7	15.3	25.8	23.6	20.4
5%	9.4	8.3	8.1	16.0	14.0	14.8
DGP P.3: Markov Regime-Switching						
10%	31.0	28.0	25.2	56.4	48.4	42.6
5%	20.6	18.2	14.8	43.0	35.8	32.0
DGP P.4: Markov Regime-Switching ARCH						
10%	35.0	31.4	27.0	62.0	53.4	49.2
5%	20.6	17.4	15.6	46.2	41.0	37.2

Notes: (i) 500 iterations and 100 bootstrap iterations for each simulation iteration;

(ii) The Bartlett kernel is used;

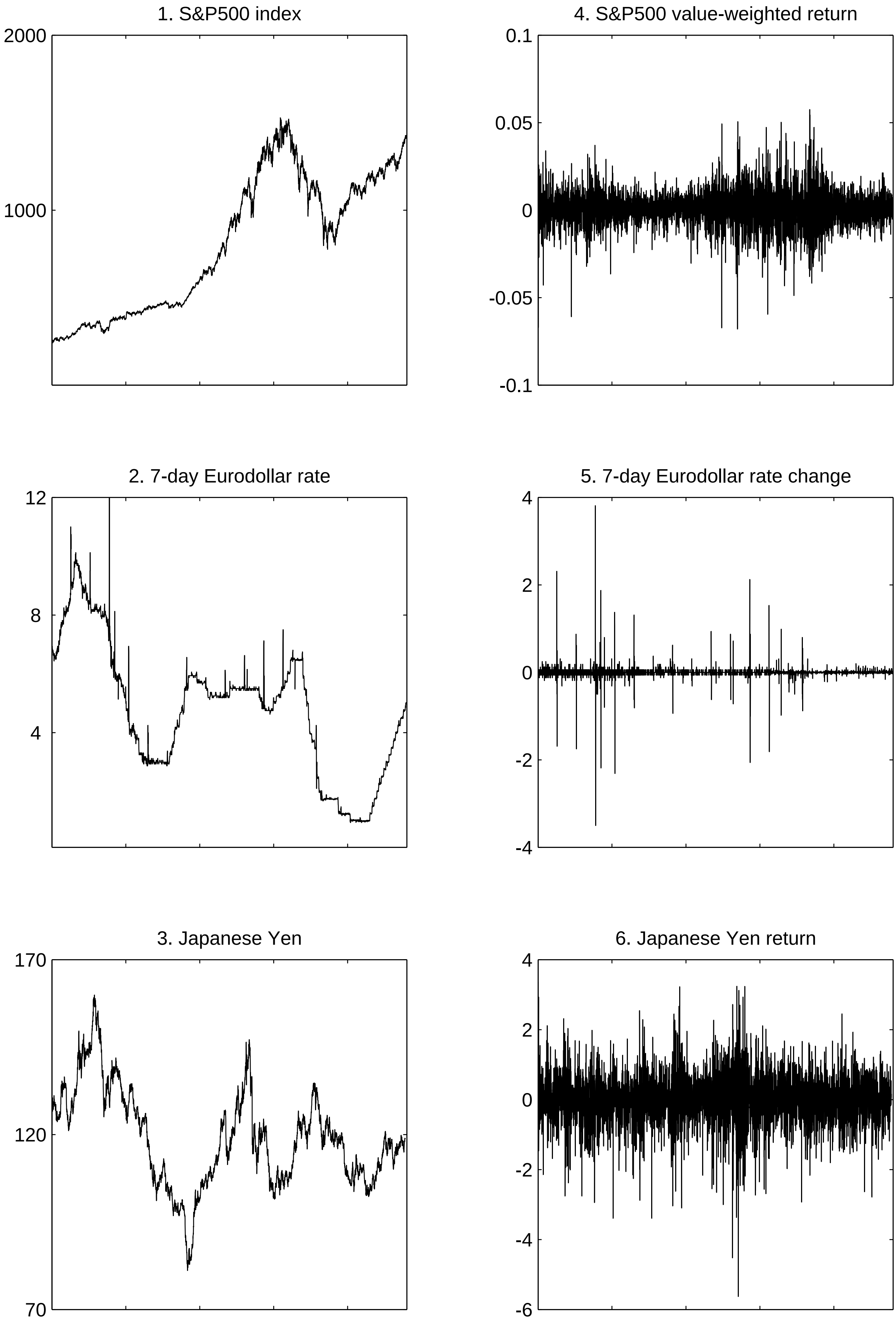
(iii) DGP S.1: $X_t = 0.5X_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim i.i.d.N\,(0,1)$; DGP S.2: $X_t = \sqrt{h_t}\varepsilon_t$, where $h_t = 0.1 + 0.1X_{t-1}^2$ and $\varepsilon_t \sim i.i.d.N\,(0,1)$;

(iv) DGP P.1: $X_t = \varepsilon_t + 0.5\varepsilon_{t-1}$, where $\varepsilon_t \sim i.i.d.N\,(0,1)$; DGP P.2: $X_t = \sqrt{h_t}\varepsilon_t$, where $h_t = 0.1 + 0.1X_{t-1}^2 + 0.8h_{t-1}$ and $\varepsilon_t \sim i.i.d.N\,(0,1)$; DGP P.3: $X_t = \begin{cases} 0.7X_{t-1} + \varepsilon_t, & \text{if } S_t = 0 \\ -0.3X_{t-1} + \varepsilon_t, & \text{if } S_t = 1 \end{cases}$, where $P\,(S_t = 1|S_{t-1} = 0) = P\,(S_t = 0|S_{t-1} = 1) = 0.9$; DGP P4: $X_t = \begin{cases} \sqrt{h_t}\varepsilon_t, & \text{if } S_t = 0 \\ 3\sqrt{h_t}\varepsilon_t, & \text{if } S_t = 1 \end{cases}$, where $h_t = 0.1 + 0.3X_{t-1}^2$, and $P\,(S_t = 1|S_{t-1} = 0) = P\,(S_t = 0|S_{t-1} = 1) = 0.9$.

Table 2 Markov test for S&P 500, interest rate and exchange rate

	01/01/1998 – 12/31/2006						01/01/1988 – 12/31/2006					
	S&P 500		7-day Eurodollar rate		Japanese Yen		S&P 500		7-day Eurodollar rate		Japanese Yen	
lag	Statistics	p-values	Statistics	p-values	Statistics	p-values	Statistics	p-values	Statistics	p-values	Statistics	p-values
10	15.04	0.0000	4.38	0.0020	1.21	0.0040	28.69	0.0000	5.74	0.0000	1.90	0.0000
11	15.81	0.0000	4.40	0.0020	1.21	0.0040	30.07	0.0000	6.07	0.0000	1.90	0.0000
12	16.51	0.0000	4.42	0.0020	1.37	0.0040	31.45	0.0000	6.36	0.0000	2.00	0.0000
13	17.19	0.0000	4.43	0.0020	1.57	0.0040	32.56	0.0000	6.62	0.0000	2.30	0.0000
14	17.83	0.0000	4.44	0.0020	1.76	0.0020	33.69	0.0000	6.84	0.0000	2.58	0.0000
15	18.43	0.0000	4.44	0.0020	1.96	0.0000	34.73	0.0000	7.02	0.0000	2.84	0.0000
16	18.97	0.0000	4.44	0.0020	2.15	0.0000	35.69	0.0000	7.16	0.0000	3.08	0.0000
17	19.46	0.0000	4.44	0.0020	2.33	0.0000	36.56	0.0000	7.29	0.0000	3.31	0.0000
18	19.90	0.0000	4.44	0.0020	2.50	0.0000	37.36	0.0000	7.39	0.0000	3.52	0.0000
19	20.29	0.0000	4.43	0.0020	2.66	0.0000	38.10	0.0000	7.48	0.0000	3.73	0.0000
20	20.65	0.0000	4.42	0.0020	2.82	0.0000	38.77	0.0000	7.55	0.0000	3.92	0.0000

Notes: (i) The bootstrap p-values are calculated by the smoothed nonparametric transition density-based bootstrap procedure described in Section 5 with 500 bootstrap iterations; (ii) The data driven lag order $\hat{p}$ is computed with the preliminary lag order ranging from 10 to 20.


1. S&P500 index
2000
1000
4. S&P500 value-weighted return
0.1
0.05
0
-0.05
-0.1
2. 7-day Eurodollar rate
12
8
4
5. 7-day Eurodollar rate change
4
2
0
-2
-4
3. Japanese Yen
170
120
70
6. Japanese Yen return
4
2
0
-2
-4
-6