# A Unified Approach to Testing Nonlinear Time Series Models

Yongmiao Hong    Yoon-Jin Lee

Cornell University & Xiamen University    Indiana University

December 2013

Yongmiao Hong, Department of Economics and Department of Statistical Science, Uris Hall, Cornell University, Ithaca, NY 14850, USA and Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen 361005, Fujian, China. Email: yh20@cornell.edu; and Yoon-Jin Lee, Department of Economics and Department of Statistics, Indiana University, Wylie Hall, Bloomington, IN 47405, USA, Email: lee243@indiana.edu.
Hong thanks supports from China National Science Foundation and the Cheung Kong Scholarship of Chinese Ministry of Education and Xiamen University. Lee thanks for financial support from Indiana University.

**ABSTRACT**

This paper proposes a unified approach to testing adequacy of nonlinear time series models. The proposed test can be applied to various nonlinear time series models, including conditional probability distribution models, Markov chain regime-switching models, conditional duration models, conditional intensity models, continuous-time jump diffusion models, continuous-time regression models, and conditional quantile and interval models. Our approach is built upon the fact that for many nonlinear time series models, model adequacy usually implies that a suitably transformed process is an independent and identically distributed (*i.i.d.*) sequence with some specified marginal distribution. Examples include the probability integral transform of an autoregressive conditional distribution model, the integrated hazard function of a conditional duration or intensity model, the time-change transform of a continuous-time regression model, and the binary transformation of an autoregressive conditional quantile or interval model. These transforms are, respectively, *i.i.d.*U[0,1], *i.i.d.*EXP(1), *i.i.d.* N(0,1) and *i.i.d.* Bernoulli($\alpha$) for some known $\alpha \in (0,1)$ when the time series models are correctly specified. The transformed process may be called the generalized residuals of a time series model since they are generalizations of Cox and Snell's (1968) concept of generalized residuals to a time series context. The proposed test checks the joint hypothesis of generalized residuals via a frequency domain approach and has omnibus power against a wide range of model misspecifications. It has a convenient null asymptotic N(0,1) distribution and is robust to dependent persistence in the underlying time series process. A Monte Carlo simulation study illustrates the merits of the approach.

*Key Words*: Binary Transformation, Generalized Residual, Generalized Spectrum, Joint Testing, Probability Integral Transform, Integrated Hazard, Time Change, Nonlinear Time Series Models.

# 1. Introduction

Nonlinear time series analysis has been advancing rather rapidly in the past thirty years (e.g., Fan and Yao 2003, Gao 2007, Granger and Teräsvirta 1993, Tjøstheim 1994, Tong 1990). A variety of nonlinear time series models have been proposed and widely used in various branches of science and social science. Unlike linear time series models, each nonlinear time series model has its own features and existing specification tests for nonlinear time series models are often model-specific. There are few unified tests in the literature that can be used to check various nonlinear time series models. In this paper, we propose a unified approach to testing various nonlinear time series models, using the fact that the adequacy of many nonlinear time series models often implies that a suitably transformed series is an independent and identically distributed (*i.i.d.*) sequence with a specified marginal distribution. For example, the probability integral transform of an autoregressive conditional distribution model, the integrated hazard function of an autoregressive conditional duration or intensity model, and the time-change transform of a continuous-time regression model (Park 2008), and the binary transform of an autoregressive conditional quantile or interval model are an *i.i.d.* sequence with a specified marginal distribution respectively when the underlying models are correctly specified. Thus, as a generally applicable approach to testing the adequacy of nonlinear time series models, one can check the joint hypothesis of the *i.i.d.* property and the specified marginal distribution of the transformed process. The transformed series may be called the generalized residual of the nonlinear time series models, since it is a generalization of Cox and Snell's (1968) generalized residual to a time series context. The transformation is essentially a filter that can capture all dynamic dependence of the time series process so that its outputs – the generalized residuals – becomes an *i.i.d.* sequence with some known distribution.

It is not a trivial task to test the joint hypothesis of the *i.i.d.* property and a specified marginal distribution for the generalized residuals. In this paper, we propose a unified approach to testing adequacy of various nonlinear time series models using the generalized spectral approach originally proposed in Hong (1999). In the present context, autocorrelation-based tests (e.g., Box and Pierce 1971) are not appropriate for nonlinear time series models because it is well-known that a time series can have zero autocorrelation but are not serially independent (e.g., an ARCH process). The idea of generalized spectrum in Hong (1999) is to first transform a time series via the characteristic function and then consider spectral analysis of the transformed series. It is an analytic tool for nonlinear time series. As an alternative to higher order spectra (Brillinger and Rosenblatt 1967a, 1967b, Subba Rao and Gabr 1984), the generalized spectrum does not require existence of any moment condition of the underlying time series process. It can capture not only the serial dependence of the underlying time series but also the shape of the marginal distribution. It is thus suitable for the aforementioned joint testing problem for the generalized residuals.

Our approach is applicable to a variety of nonlinear time series models. It can test such popular time series models as autoregressive conditional density models (e.g., Hansen 1994), Markov chain regime switching models, autoregressive conditional duration models (Engle and Russell 1998), autoregressive conditional intensity model (Russell 1999), continuous time jump diffusion model (e.g., Barndorff-Nielsen and Shephard 2001), continuous time regression models (Park 2008), and autoregressive conditional quantile and interval models (e.g., Koenker and Xiao 2006). We allow but do not restrict to test location-scale time series models that capture all serial dependence of the time series process by the first two conditional moments. Our approach is applicable to time series models with either continuous or discrete distributions. Another important feature of the proposed test is its robustness to persistence in the original time series process, thanks to the fact that the generalized residuals are always *i.i.d.* under the null hypothesis. This is appealing for many applications since, for example, most high-frequency economics and financial time series have highly persistent dependence. It is well known that statistical inference procedures often do not perform well in finite samples when the underlying time series process is highly persistent. The robustness of the size performance of our procedure avoids the use of bootstrap methods which usually involve reestimation of nonlinear time series models and are computationally costly. For example, the likelihood surface of Markov chain regime-switching models is often found to be highly irregular and contains several local maxima, and so it is hard to achieve a convergence for parameter estimation. On the other hand, the proposed test does not have to formulate an alternative model and has a null asymptotic N(0,1) distribution. Moreover, the sampling variation of parameter estimation uncertainty has no impact on the asymptotic normal distribution of the proposed test statistic. Thus, there is no need to calculate otherwise tedious delta expressions of a nonlinear time series model. These features lead to a convenient inference procedure.

Section 2 introduces hypotheses of interest and provides motivation. The generalized spectral density based test statistics are given in Section 3. Section 4 derives the asymptotic normal distribution of the proposed tests and investigates their asymptotic power property. Section 5 examines their finite sample performance via Monte Carlo experiments. Section 6 concludes. All mathematical proofs are collected in the appendix. Throughout, we denote $C$ for a generic bounded constant, $A^*$ for the complex conjugate of $A$, $\operatorname{Re} A$ for the real part of $A$, and $||A||$ for the Euclidean norm of $A$. All limits are taken as the sample size $T \to \infty$. The GAUSS code to implement our tests is available from the authors upon request.

## 2. Hypothesis of Interest and Literature

### *2.1 Hypothesis of Interest and Motivation*

In time series analysis, one is often interested in modeling the dynamics of a time series process $\{Y_t\}$. Suppose a parametric nonlinear model, say $\mathcal{M}(\theta)$, is used to capture the dynamics of $\{Y_t\}$, where $\theta$ is an unknown finite-dimensional parameter vector to be estimated using observed data. We are interested in proposing a generally applicable method to check the adequacy of the model $\mathcal{M}(\theta)$.

Residual-based testing has been a popular approach in time series analysis. In linear time series analysis, for example, Box and Pierce (1971) proposed a portmanteau test based on the estimated residuals of a linear ARMA model. In nonlinear time series modeling, the concept of residuals is not so obvious, but we can make use of the concept of generalized residuals in spirit of Cox and Snell (1968). For many nonlinear time series models, there exists some transformation or filter that can capture all dependence structure of the underlying time series so that its outputs become an *i.i.d.* sequence with some known distribution. Specifically, given observed data $\{Y_t, I_{t-1}\}_{t=1}^T$, where $Y_t$ is a real-valued dependent variable and $I_{t-1}$ may contain lagged dependent variables and lagged exogenous variables $X_t$, we can define $Z_t(\theta_0) \equiv H(Y_t, I_{t-1}, \theta_0)$ given by a known measurable transformation $H$ and an unknown parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$. The transformed series $\{Z_t(\theta_0)\}$ can be called the generalized residuals of the nonlinear time series models since they are the generalizations of Cox and Snell's (1968) concept of generalized residuals to a time series context. To illustrate this concept and the scope of our approach, we consider a variety of nonlinear time series models below.

Example 1 [GARCH And Nonnegative Process]:

The GARCH model has been one of the popular nonlinear time series models:

$$\begin{cases} Y_t = \mu_t + \sqrt{h_t}\varepsilon_t, \\ \mu_t = \mu(I_{t-1}, \theta), \\ h_t = h(I_{t-1}, \theta), \end{cases} \tag{2.1}$$

where $\mu(I_{t-1}, \theta)$ and $h(I_{t-1}, \theta)$ are parametric models for $E(Y_t|I_{t-1})$ and $Var(Y_t|I_{t-1})$ respectively, and $I_{t-1}$ is the information set available at time $t-1$ which is the $\sigma$-field generated by the past history of $Y_t$, $\{Y_s, s < t\}$. We allow for an infinite past history of information, i.e., we allow but do not assume that $Y_t$ is Markov. Suppose further the standardized innovation $\{\varepsilon_t\}$ is specified to follow some conditional distribution $g(\varepsilon|I_{t-1}, \theta)$. Then the conditional density model of $Y_t$ given $I_{t-1}$ is

$$f(y|I_{t-1}, \theta) = \frac{1}{\sqrt{h_t}} g\left(\left.\frac{y-\mu_t}{\sqrt{h_t}}\right| I_{t-1}, \theta\right), \quad -\infty < y < \infty.$$

Define the dynamic probability integral transform

$$Z_t(\theta) = \int_{-\infty}^{Y_t} f(y|I_{t-1}, \theta)dy. \tag{2.2}$$

Then $Z_t(\theta_0)$ is $i.i.d. U[0,1]$ at some parameter value $\theta_0$ when $f(y|I_{t-1}, \theta)$ is correctly specified, i.e., when $f(y|I_{t-1}, \theta_0)$ coincides with the true conditional probability density of $Y_t$ given $I_{t-1}$. See Rosenblatt (1952). This probability integral transform is called the generalized residual of the GARCH model in (2.1). Note that $Z_t(\theta_0)$ is always *i.i.d.* whereas the standardized innovation $\{\varepsilon_t\}$ may not be *i.i.d.* even when the model $f(y|I_{t-1}, \theta)$ is correctly specified. One example is Hansen's (1994) autoregressive conditional density model, which allows parametric specifications for conditional dependence beyond the mean and variance. Specifically, Hansen (1994) assumes

$$f(y|I_{t-1},\theta) = f_0\left[y|\alpha(I_{t-1},\theta)\right],$$

where $f_0(y|\cdot)$ is a generalized skewed $t$-distribution and $\alpha(I_{t-1},\theta)$ is a low-dimensional vector that characterizes the first four time-varying conditional moments of $Y_t$ given $I_{t-1}$.

Likewise, we can also consider a general framework of multiplicative error models for nonnegative time series processes, which are common in practice. For example, one could model the volume of shares over a 10-minute period, or the high price minus the low price over a time period or the ask price minus the bid price, or the time between trades, or the number of trades in a period (Engle 2002). The models for nonnegative processes can be modeled as a multiplicative form similar to the GARCH structure:

$$\begin{cases} Y_t = \psi_t \varepsilon_t, \\ \psi_t = \psi(I_{t-1},\theta), \end{cases} \tag{2.3}$$

where $\psi_t$ is a parametric model for $E(Y_t|I_{t-1})$, $\varepsilon_t$ is a multiplicative error with $E(\varepsilon_t|I_{t-1}) = 1$ when $\psi_t$ is correctly specified for $E(Y_t|I_{t-1})$. One example of multiplicative error models is the autoregressive conditional duration models proposed by Engle and Russell (1998) and Engle (2000), where $Y_t$ is the arrival time intervals between consecutive events such as the occurrence of a trade or a bid-ask quote.

The dynamic probability integral transform is also applicable to many other time series models, including Markov chain regime switching models and continuous-time jump diffusion models, as illustrated below.

Example 2 [Markov Chain Regime-Switching model]:

The Markov chain regime-switching model has been popularly used in time series econometrics (e.g., Hamilton 1994, Ch.22). It posits that the conditional distribution of a time series depends on an underlying latent state, which can take one of a finite number of values and evolves through time as a Markov chain. This model allows for complex nonlinear dynamics and yet remains tractable. Testing regime-switching models has been an interesting problem in time series and yet little effort has been devoted to specification testing for this class of models. In fact, the generalized residual provides a convenient way to test this class of models. Consider a time series $Y_t$, the conditional distribution of which depends on the latent state variable $S_t$, which occurs at time $t$ and takes $K$ discrete values indexed by $j \in \{1,\cdots,K\}$. Assume that the state dependent conditional distribution of $Y_t$ is given as follows:

$$f\left(y|S_t = j, I_{t-1}\right) = f_0\left(y|S_t = j, I_{t-1},\theta_0\right) \text{ for some } \theta_0 \in \Theta,$$

where $f_0(y|\cdot,\cdot)$ is a known parametric density, $I_{t-1}$ denotes the information set available in period $t-1$, and the latent regime $S_t$ evolves through time as a first order Markov chain with transition probabilities given by the $K \times K$ transition matrix $P$ with element $(i,j)$

$$P(S_t = j | S_{t-1} = i) = p_{ij}, \quad i, j = 1, \cdots, K.$$

Then when the Markov chain regime-switching model is correctly specified, we have

$$Z_t(\theta_0) = \sum_{j=1}^{K} P(S_t = j | I_{t-1}) \int_{-\infty}^{Y_t} f_0(y | S_t = j, I_{t-1}, \theta_0) dy \sim i.i.d.U[0, 1].$$

Example 3 [Continuous-time Jump diffusion models]:

Continuous-time jump diffusion models have been popularly used in mathematical finance. Consider a class of jump diffusion models (e.g., Barndorff-Nielsen and Shephard 2001, Duffie, Pan and Singleton 2001):

$$dY_t = \mu(Y_t, \theta)dt + \sigma(Y_t, \theta)dW_t + JdN_t,$$

where $\mu(Y_t, \theta)$ is a drift model, $\sigma(Y_t, \theta)$ is a diffusion model, $W_t$ is a Brownian motion and $N_t$ is a Poisson process, which determines the random arrival of jump $J$, with the intensity $\lambda_t(\theta)$. Given the specifications of $\mu(Y_t, \theta), \sigma(Y_t, \theta)$ and $\lambda_t(\theta)$, the transition density model of $Y_t$ is then determined as a parametric model $f(y | Y_{t-\Delta}, \theta)$, where $\Delta$ is any given sampling frequency. When the jump diffusion model is correctly specified, we have

$$Z_t(\theta_0) = \int_{-\infty}^{Y_t} f(y | Y_{t-\Delta}, \theta_0) dy \sim i.i.d.U[0, 1].$$

This result holds when the Brownian motion $W_t$ is replaced with the more general Levy process.

In addition to the probability integral transform, there are alternative transforms which can be used to construct the generalized residuals. In duration or survival analysis, for example, the integrated hazard function will follow an $i.i.d.EXP(1)$ when an autoregressive conditional duration or survival model is correctly specified.

Example 4 [Autoregressive Conditional Intensity model]:

Define a counting process $N_t = \{N(t), t \geq 0\}$ by $N(t) = \sum_{i=1}^{\infty} \mathbf{1}(T_i \leq t)$ for all $t \geq 0$, where $\mathbf{1}(\cdot)$ is the indicator function. The corresponding point process is the random arrival time $\{T_i, i = 0, 1, 2, \cdots\}$. Then, the conditional intensity function (hazard function), which assesses the instantaneous risk of demise (e.g., credit default) at time $t$, is given as follows:

$$\lambda(t | \mathcal{F}_t) = \lim_{\Delta t \to 0} \frac{\Pr(N(t + \Delta t) - N(t) > 0 | \mathcal{F}_t)}{\Delta t},$$

where $\mathcal{F}_t = \sigma(N_s, s \in [0, t])$ is the history generated by $N_t$, and $N(t)$ is assumed to be adapted to the filtration $\mathcal{F}_t$. In time series survival analysis, a parametric model $\lambda_0(t | \mathcal{F}_t, \theta)$ is often used to approximate the hazard function $\lambda_0(t | \mathcal{F}_t)$. We can define the generalized residuals of the hazard model

$\lambda(t|\mathcal{F}_t)$ by

$$Z_i(\theta) = 1 - \int_0^t \lambda_0(s|\mathcal{F}_i, \theta) ds.$$

Then $Z_i(\theta_0) \sim i.i.d.EXP(1)$ when the hazard model $\lambda_0(t|\mathcal{F}_t, \theta)$ is correctly specified (Yashin and Arjas (1988)).

We can also define generalized residuals $\{Z_t(\theta)\}$ for an instantaneous conditional mean model in continuous time. The following time-change transform for the continuous-time regression model (Park 2008) provides such an example.

Example 5 [Park's (2008) time change in continuous-time]:

Consider a continuous time regression model (Park 2008)

$$dY_t = \mu(Y_t, \theta) dt + dU_t,$$

where $\{Y_t\}$ is a stochastic process, $\{\mathcal{F}_t\}$ is a filtration to which $\{Y_t\}$ is adapted, $\mu(Y_t, \theta)$ is a parametric model for the instantaneous conditional mean $\lim_{\delta \to 0^+} E\left[ \frac{Y_{t+\delta} - Y_t}{\delta} \middle| \mathcal{F}_t \right]$, and $\{U_t\}$ is a martingale process with respect to the filtration $\{\mathcal{F}_t\}$ so that $dU_t$ is a martingale difference sequence (*m.d.s.*) with $E(dU_t|\mathcal{F}_t) = 0$. Define a time change, a non-decreasing collection of stopping times, by

$$T_t = \inf_{s>0} \{\langle U \rangle_s > t\},$$

where $\langle U \rangle_t$ is the quadratic variation process of $U_t$. Then we have

$$U_{T_t} = V_t \quad \text{or} \quad U_t = V_{\langle U \rangle_t},$$

where $V_t$ is the standard Brownian motion (see Park 2008). Thus, with an appropriate time change, the martingale regression (instantaneous conditional mean) model can always be transformed into a regression model with the error process given by the Brownian motion. It follows that with the time change $T_t$, we have

$$dY_{T_t} = \mu(Y_{T_t}, \theta) dT_t + dU_{T_t} = \mu(Y_{T_t}, \theta) dT_t + dV_t,$$

and $V_t \equiv V_t(\theta)$ is a standard Brownian motion at $\theta = \theta_0$ when the instantaneous conditional mean model $\mu(Y_t, \theta)$ is correctly specified. This imples that the error process in the time changed regression model is the standard Brownian motion. We can thus define a generalized residual for the continuous-time regression model as

$$\begin{aligned} Z_t(\theta) &= V_t(\theta) - V_{t-\Delta}(\theta) \\ &= \Delta^{-1/2} \left( Y_{T_{t\Delta}} - Y_{T_{(t-1)\Delta}} - \int_{T_{(t-1)\Delta}}^{T_{t\Delta}} \mu(Y_t, \theta) dt \right), \quad t = 1, \cdots, n, \end{aligned}$$

where $\Delta$ is any given sampling frequency for the observed data. Then $Z_t(\theta_0)$ is $i.i.d.N(0,1)$ for some

$\theta_0 \in \Theta$ when the instantaneous conditional mean model $\mu(Y_t, \theta)$ is correctly specified. Therefore, the method based on generalized residuals can serve as a specification test for continuous-time models.

The last example is the class of autoregressive conditional quantile and interval models.

Example 6 [Autoregressive Conditional Quantile and Interval Models]:

An autoregressive conditional $\alpha$-quantile model $Q_\alpha(I_{t-1}, \theta)$ obeys the following condition

$$P[Y_t \leq Q_\alpha(I_{t-1}, \theta)|I_{t-1}] = \alpha \text{ for some } \theta = \theta_0$$

when the quantile model is correctly specified. Examples are J.P. Morgan's RiskMetrics and Engle and Manganelli's (2004) CAViaR models for Value at Risk in financial risk management. Also see Koenker and Xiao's (2006) quantile autoregression model.

We define the binary stochastic process

$$Z_t(\theta) = \mathbf{1}\left[Y_t \leq Q_\alpha(I_{t-1}, \theta)\right].$$

Under correct model specification of $Q_\alpha(I_{t-1}, \theta)$, $\{Z_t(\theta_0)\}$ is an *i.i.d.* Bernoulli($\alpha$) sequence, and this property can be used to test adequacy of autoregressive conditional quantile models.

This method also applies to autoregressive confidence interval models. $100(1-\alpha)\%$ interval forecast (e.g., Christoffersen (1998)) $(L_\alpha(I_{t-1}, \theta), U_\alpha(I_{t-1}, \theta))$ is correctly specified iff

$$P\left[L_\alpha(I_{t-1}, \theta) \leq Y_t \leq U_\alpha(I_{t-1}, \theta)|I_{t-1}\right] = 1 - \alpha \quad a.s.$$

where $L_\alpha(I_{t-1}, \theta)$ and $U_\alpha(I_{t-1}, \theta)$ are lower and upper bounds of the conditional interval model for $Y_t$ given $I_{t-1}$ at confidence level $1-\alpha$. Define

$$Z_t(\theta) = \mathbf{1}\left[L_\alpha(I_{t-1}, \theta) \leq Y_t \leq U_\alpha(I_{t-1}, \theta)\right],$$

then under correct model specification of $(L_\alpha(I_{t-1}, \theta), U_\alpha(I_{t-1}, \theta))$,

$$\{Z_t(\theta_0)\} \sim i.i.d. \ Bernoulli(\alpha) \text{ sequence.}$$

All the aforementioned examples can be formulated as a unified hypothesis of interest

$$\mathbb{H}_0 : \{Z_t(\theta_0)\} \sim i.i.d. F_{\theta_0}(z) \text{ for some unknown } \theta_0 \in \Theta, \tag{2.4}$$

where $F_\theta(\cdot)$ is a known probability distribution function, which can be a continuous, discrete, or mixed continuous and discrete distribution, and parameter $\theta$ is unknown. This provides a unified approach to testing various nonlinear time series models, as illustrated above. There are other advantages of testing $\mathbb{H}_0$ via the generalized residual $Z_t(\theta)$. For example, given the *i.i.d.* property of $\{Z_t(\theta)\}$ under $\mathbb{H}_0$, the size of the test is expected to be robust to dependence persistence of $\{Y_t\}$ in finite samples.

Intuitively, in some cases (i.e., location-scale time series models which capture all serial dependence via the first two conditional moments), the *i.i.d.* property of $\{Z_t(\theta)\}$ characterizes correct specification of the dynamic dependence structure of a time series model, and the specified parametric distribution $F_\theta(z)$ characterizes correct specification of the marginal error distribution of the time series model. The goal of this paper is to propose a novel and generally applicable test to various nonlinear time series models which can be characterized as $\mathbb{H}_0$ when they are correctly specified.

The difficulty of testing $\mathbb{H}_0$ in (2.4) is this joint hypothesis in a nonlinear time series setup. Often, the Kolmogorov-Smirnov (KS) test is suggested to test $\mathbb{H}_0$. However, the KS test only focuses on the marginal distribution $F_\theta(\cdot)$ and does not check the serial dependence structure in $\{Z_t(\theta)\}$. It has no power if $Z_t(\theta)$ follows a marginal distribution $F_\theta(\cdot)$ but $\{Z_t(\theta)\}$ is serially dependent. Moreover, the asymptotic critical values of the KS statistic will be affected by sampling variation of parameter estimation.

*2.2 Comparison with the Literature*

There have been tests for some specific nonlinear time series models using the probability integral transforms. For example, Bai (2003) proposes a generalized KS test using the probability integral transform and Khmaladze's (1981) martingale transform. The latter nicely removes the impact of parameter estimation uncertainty, delivering an asymptotically distribution-free test. However, Bai's (2003) test checks $U[0,1]$ under the *i.i.d.* property of the generalized residuals, and it still has no power if $Z_t(\theta)$ is $U[0,1]$ but not serially independent.

Thompson (2008) uses probability integral transforms to test continuous-time diffusion models. He uses the Cramer-von Mises-type statistic based on the empirical distribution function of $Z_t(\theta)$ (to test uniformity) and the periodogram (to test independence). This test checks the joint hypothesis of *i.i.d.* $U[0,1]$ but it may lack power against nonlinear alternatives because periodogram will miss dependence with zero autocorrelation. Furthermore, sampling variation of the parameter estimation has impact on the asymptotic distribution which calls for use of a parametric bootstrap. Also in the continuous time context, Hong and Li (2005) test both *i.i.d.* and $U[0,1]$ simultaneously by using a smoothed bivariate nonparametric kernel density estimator of the probability integral transforms, but they check individual lags rather than all lags jointly and nonparametric smoothing is required at each individual lag.

Berkowitz (2001) considers a test for density forecast evaluation by extending the probability integral transform to normality (i.e., using quantile transformation). That is, the probability integral transform is further converted into $i.i.d. N(0,1)$ by inverting the CDF using the normal distribution. Based on normality, Berkowitz (2001) considers a likelihood-ratio test. However, the LR test only has power to detect nonnormality through the first two moments of the distribution, and need to specify the likelihood of alternative. Specifically, he considers the combined statistic for the joint test of independence and zero mean and unit variance, and only considers an AR(1) alternative. Also, he does not consider the impact of parameter estimation by noting that in his context "the cost of abstracting from parameter uncertainty may not be severe".

Hong and T.Lee (2003) propose a test for the following class of location-scale nonlinear time series models

$$Y_t = \mu(I_{t-1}, \theta) + \sqrt{h(I_{t-1}, \theta)}\varepsilon_t. \tag{2.5}$$

where the standardized innovation $\{\varepsilon_t\}$ is an *i.i.d.* sequence. For this class of time series models, the first two conditional moments capture all serial dependence of $Y_t$. Hong and T.Lee (2003) check adequacy of model (2.5) by testing whether the standardized innovations $\{\varepsilon_t\}$ is an *i.i.d.* sequence. This test does not apply to test $\mathbb{H}_0$ here, because it only checks serial dependence and does not check the marginal distribution $F_\theta(z)$. In other words, Hong and T. Lee's (2003) test, even when applied to the generalized residuals $\{\hat{Z}_t\}$, will have no power when $Z_t$ is *i.i.d.* but its marginal distribution is not $F_\theta(\cdot)$. This can occur, e.g., when $\{Y_t\}$ is a GARCH(1,1) process with an *i.i.d.* $t$-distributed innovation but it is specified as a GARCH(1,1) model with an *i.i.d.* N(0,1) innovation. In this case, the probability integral transforms are an *i.i.d.* sequence but with a non-uniform distribution. As a result, Hong and T.Lee's (2003) test has no power despite the misspecification in the distribution of $\varepsilon_t$. Note that each generalized covariance function $\sigma_j(u, v)$ $\forall j$ in our approach is different from one in Hong and T. Lee (2003). Each $\sigma_j(u, v)$ contains the information about the marginal distribution as well as serial dependence. Thus, our approach is fundamentally different from Hong and T. Lee (2003).

On the other hand, there exist time series models in the form of (2.5) but the standardized innovation $\{\varepsilon_t\}$ is not *i.i.d.* An example is Hansen's (1994) autoregressive conditional density model. For this model, $\{\varepsilon_t\}$ is a conditionally homoskedastic *m.d.s.* with $E(\varepsilon_t|I_{t-1}) = 0$ and $Var(\varepsilon_t|I_{t-1}) = 1$ but its conditional higher order moments (e.g., skewness and kurtosis) are time-varying even when the time series model is correctly specified. Hong and T.Lee's (2003) test cannot be applied to test this model, because the *i.i.d.* property of $\{\varepsilon_t\}$ is not a characteristic of the correct specification of these models. Of course, Hong and T. Lee's (2003) test can be applied to the generalized residuals $\{Z_t\}$ to test serial dependence of $\{Z_t\}$, but as pointed out above, it ignores testing the marginal distribution $F_\theta(z)$ of $\mathbb{H}_0$.

In this paper, we will propose a new generally applicable test for $\mathbb{H}_0$ that avoid the drawback of the aforementioned tests for various nonlinear time series models. We test the *i.i.d.* property and $F_\theta(\cdot)$ jointly using a generalized spectral approach. The test statistic has an asymptotic N(0,1) distribution under $\mathbb{H}_0$ and parameter estimation has no impact on the limit distribution, thus resulting in a convenient procedure. Also, we test many lags simultaneously, and our approach naturally provides a downward weighting, which may enhance power when a large number of lag orders are considered.

## 3. Approach and Test Statistics

To describe our approach, we first introduce the generalized spectrum proposed in Hong (1999).

For notational simplicity, we put $Z_t = Z_t(\theta_0)$, where $\theta_0 = p\lim(\hat{\theta})$ and $\hat{\theta}$ is a parameter estimator for the parametric model $\mathcal{M}(\theta)$ of time series process $\{Y_t\}$. Suppose $\{Z_t\}$ is a strictly stationary time series process. Then, following Hong (1999), the generalized spectral density of $\{Z_t\}$ is defined by

$$f(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v) e^{-ij\omega}, \quad i = \sqrt{-1}, -\infty < u, v < \infty, \tag{3.1}$$

where $\sigma_j(u, v)$ is a generalized autocovariance function

$$\sigma_j(u, v) = \text{cov}(e^{iuZ_t}, e^{ivZ_{t-|j|}}), \quad -\infty < u, v < \infty. \tag{3.2}$$

Intuitively, the generalized spectrum $f(\omega, u, v)$ is the spectrum of the transformed time series $\{e^{iuZ_t}\}$. As the power spectral density, which is the Fourier transform of the autocovariance function of $\{Z_t\}$, is a basic analytic tool for linear time series, the generalized spectral density $f(\omega, u, v)$ is a basic analytic tool for nonlinear time series and an alternative to higher order spectra (Brillinger and Rosenblatt 1967a, 1967b, Subba Rao and Gabr 1984). The exponential function transformation enables $f(\omega, u, v)$ to capture both linear and nonlinear dependence. Observe that

$$\sigma_j(u, v) = \text{cov}(e^{iuZ_t}, e^{ivZ_{t-|j|}}) = \varphi_j(u, v) - \varphi(u)\varphi(v),$$

where $\varphi_j(u, v) = Ee^{iuZ_t + ivZ_{t-|j|}}$ is the joint characteristic function of the pair of random variables $(Z_t, Z_{t-|j|})$, and $\varphi(u) = Ee^{iuZ_t}$ is the marginal characteristic function of $Z_t$. The function $\sigma_j(u, v) = 0$ for all $u, v$ if and only if $Z_t$ and $Z_{t-|j|}$ are independent. Thus, $f(\omega, u, v)$ can capture any pairwise serial dependence, including the serial dependence patterns for which $\{Z_t\}$ is serially uncorrelated but not serially independent (e.g., an ARCH process). This can be also seen by taking a Taylor series expansion of $f(\omega, u, v)$ with respect to $(u, v)$ at the origin $(0, 0)$ :

$$f(\omega, u, v) = \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \frac{(iu)^m (iv)^l}{m!l!} f^{(0,m,l)}(\omega, 0, 0), \tag{3.3}$$

where the derivative function

$$f^{(0,m,l)}(\omega, 0, 0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \text{cov}(Z_t^m, Z_{t-|j|}^l) e^{-ij\omega}.$$

Note that the partial derivative $f^{(0,1,1)}(\omega, 0, 0)$ is the conventional power spectral density of $\{Z_t\}$. The Taylor series expansion in (3.3) requires that all moments of $Z_t$ exist. The generalized spectrum in (3.1) does not have such a requirement. It is well defined even when the moment of $Z_t$ does not exist. Moreover, since the generalized autocovariance function $\sigma_j(u, v)$ is defined in terms of the characteristic function, $f(\omega, u, v)$ can also be used to capture the marginal and pairwise distributional properties of the time series process $\{Z_t\}$. Thus, it provides a natural approach to testing the joint hypothesis $\mathbb{H}_0$. Our idea here is to compare the shapes of the generalized spectral density function under $\mathbb{H}_0$ and under the alternative to $\mathbb{H}_0$ respectively. When $\mathbb{H}_0$ holds, the generalized autocovariance function $\sigma_j(u, v)$ becomes the following:

$$\begin{aligned}\sigma_j(u,v) &= \sigma_j(u,v|\theta_0) \\ &= \text{Cov}_{\theta_0}(e^{iuZ_t}, e^{ivZ_{t-j}}) \\ &= \begin{cases} \varphi_{\theta_0}(u+v) - \varphi_{\theta_0}(u)\varphi_{\theta_0}(v) & \text{if } j = 0 \\ 0 & \text{if } j \neq 0 \end{cases}\end{aligned} \tag{3.4}$$

where $Cov_{\theta_0}(\cdot, \cdot)$ is the covariance operator under $\mathbb{H}_0$ and

$$\varphi_{\theta_0}(u) = E_{\theta_0}(e^{iuZ_t}) = \int e^{iuz} dF_{\theta_0}(z) \tag{3.5}$$

is the marginal characteristic function of the distribution $F_{\theta_0}(z)$. It follows that under $\mathbb{H}_0$, $f(\omega, u, v)$ becomes "a flat spectrum":

$$f_{\theta_0}(\omega, u, v) \equiv \frac{1}{2\pi}\sigma_0(u, v|\theta_0), \tag{3.6}$$

which is a constant function of frequency $\omega \in [-\pi, \pi]$. This may be called the restricted generalized spectral density implied by $\mathbb{H}_0$. It can be consistently estimated by

$$f_{\hat{\theta}}(\omega, u, v) = \frac{1}{2\pi}\sigma_0(u, v|\hat{\theta}), \quad \omega \in [-\pi, \pi], \tag{3.7}$$

where $\hat{\theta}$ is a consistent estimator for $\theta_0$ under $\mathbb{H}_0$.

Under the alternative to $\mathbb{H}_0$, the generalized spectral density $f(\omega, u, v)$ can be consistently estimated by a smoothed kernel estimator

$$\hat{f}(\omega, u, v) = \frac{1}{2\pi}\sum_{j=1-T}^{T-1} k(j/p)(1 - |j|/T)\hat{\sigma}_j(u, v)e^{-ij\omega}, \tag{3.8}$$

where

$$\hat{\sigma}_j(u, v) = \frac{1}{T - |j|}\sum_{t=|j|+1}^{T} \hat{\varphi}_t(u)\hat{\varphi}_{t-j}(v), \tag{3.9}$$

$\hat{\varphi}_t(u) = e^{iu\hat{Z}_t} - \hat{\varphi}(u)$, $\hat{\varphi}(u) = T^{-1}\sum_{t=1}^{T} e^{iu\hat{Z}_t}$, $\hat{Z}_t = Z_t(\hat{\theta})$ is the estimated generalized residual of the time series model $\mathcal{M}(\theta)$, $k(\cdot)$ is a kernel function that assigns weights to various lag orders, and $p = p(T)$ is a bandwidth. The estimator $\hat{f}(\omega, u, v)$ may be called an unrestricted generalized spectral density estimator. Under $\mathbb{H}_0$, this unrestricted estimator will converge to the same limit as the restricted estimator $f_{\hat{\theta}}(\omega, u, v)$. If they converge to different limit functions, the null hypothesis $\mathbb{H}_0$ is rejected.

To obtain a global measure of the discrepancy between two function-valued estimators $\hat{f}(\omega, u, v)$

and $f_{\hat{\theta}}(\omega, u, v)$, we use the quadratic form

$$\begin{aligned}\hat{Q} &= \pi T \iiint_{-\pi}^{\pi} \left| \hat{f}(\omega, u, v) - f_{\hat{\theta}}(\omega, u, v) \right|^2 d\omega dW(u) dW(v) \\ &= \frac{1}{2} \iint T \left| \hat{\sigma}_0(u, v) - \sigma_0(u, v|\hat{\theta}) \right|^2 dW(u) dW(v) + 2 \sum_{j=1}^{T-1} k^2(j/p)(T-j) \iint \left| \hat{\sigma}_j(u, v) \right|^2 dW(u) dW(v) \\ &= \sum_{j=1-T}^{T-1} a_T(j)(T-j) \iint \left| \tilde{\sigma}_j(u, v) - \sigma_j(u, v|\theta_0) \right|^2 dW(u) dW(v) + O_P(1) \text{ under } \mathbb{H}_0, \qquad (3.10)\end{aligned}$$

where

$$a_T(j) = \begin{cases} \frac{1}{2} k^2(j/p) & \text{if } j = 0 \\ k^2(j/p) & \text{if } j \neq 0 \end{cases} = \begin{cases} \frac{1}{2} & \text{if } j = 0 \\ k^2(j/p) & \text{if } j \neq 0, \end{cases}$$

and $\tilde{\sigma}_j(u, v)$ be defined in the same way as $\hat{\sigma}_j(u, v)$ with the unobservable generalized residuals $\{Z_t \equiv Z(I_{t-1}, \theta_0)\}_{t=1}^T$. The last equality in (3.10) holds only under $\mathbb{H}_0$ because $\sigma_j(u, v|\theta_0) = 0$ for all $j \neq 0$ under $\mathbb{H}_0$, and replacing $\hat{\theta}$ with $\theta_0$ results in an effect of $O_P(1)$.

Intuitively, the term associated with $j = 0$ in (3.10) checks whether $\{Z_t\}$ has a marginal distribution $F_{\theta_0}(z)$, and the terms associated with all nonzero lags in (3.10) check whether $\{Z_t\}$ is serially independent. Thus, the generalized spectral approach can be used to test the joint hypothesis $\mathbb{H}_0$. It has the advantage over the KS-type test in the sense that the latter only tests the marginal distribution specification and does not test the serial dependence of $\{Z_t\}$.

Our spectral kernel estimation approach provides a flexible weighting for various lags via the kernel function. For many commonly used kernels, $k(\cdot)$ is downward weighting (see, e.g., Priestley 1981, p.442). Thus, the term with $j = 0$, which focuses on the marginal distribution, receives a largest weight $a_T(0) = \frac{1}{2}$, while other terms with $j \neq 0$ receives a downward weighting scheme $k^2(j/p)$ as $j$ increases. Downward weighting is expected to be more powerful than uniform weighting against many alternatives of practical importance as the state today is more affected by the recent events than the remote past events. When a large $p$ is employed to capture serial dependence at higher order lags, downward weighting will alleviate the loss of a large number of degrees of freedom, thus enhancing good power of the proposed test.

The proposed test statistic is an appropriately centered and scaled version of the quadratic form $\hat{Q}$ :

$$\hat{M}_1 = \frac{\hat{Q} - \hat{A}}{\sqrt{\hat{V}}}, \qquad (3.11)$$

where

$$\hat{A} = \int \left[ T^{-1} \sum_{t=1}^{T} |\hat{\varphi}_t(u) \hat{\varphi}_t(v)|^2 - \left| \sigma_0(u, v|\hat{\theta}) \right|^2 \right] dW(u) dW(v) + 2 \left[ \int \sigma_0(u, -u|\hat{\theta}) dW(u) \right]^2 \sum_{j=1}^{T-1} a_T(j),$$

$$\hat{V} = \left[ \iint |\sigma_0(u, v|\hat{\theta})|^2 dW(u) dW(v) \right]^2 2 \sum_{j=1-T}^{T-1} a_T^2(j).$$

The factors $\hat{A}$ and $\hat{V}$ are approximately the mean and variance of the quadratic form $\hat{Q}$ under $\mathbb{H}_0$. Since the centering factor $\hat{A}$ is a bit tedious to compute, we also define the following simplified and asymptotically equivalent test statistic:

$$\hat{M}_2 = \frac{\hat{Q} - \left[ \int \sigma_0(u, -u|\hat{\theta}) dW(u) \right]^2 \sum_{j=1-T}^{T-1} a_T(j)}{\sqrt{\left[ \iint |\sigma_0(u, v|\hat{\theta})|^2 dW(u) dW(v) \right]^2 2 \sum_{j=1-T}^{T-1} a_T^2(j)}}. \tag{3.12}$$

Both $\hat{M}_1$ and $\hat{M}_2$ are asymptotically $N(0, 1)$ under $\mathbb{H}_0$ as $T \to \infty$ (see Theorem 1 below). Since these tests are the centered and scaled versions of the quadratic form $\hat{Q}$, the asymptotic normality implies that a properly scaled version of the $\hat{Q}$ statistic is asymptotically Chi-squared distributed with a large numbers of degrees of freedom. For example, under $\mathbb{H}_0$, we have that when $T \to \infty$,

$$\frac{2\hat{A}}{\hat{V}} \hat{Q} \overset{a}{\sim} \chi^2_{\hat{q}}, \tag{3.13}$$

where the degrees of freedom

$$\hat{q} = \frac{2\hat{A}^2}{\hat{V}} = q[1 + o_P(1)] \equiv \frac{4 \left[ \int \sigma_0(u, -u|\theta_0) dW(u) \right]^4 \left[ \int k^2(z) dz \right]^2}{\left[ \iint \sigma_0(u, v|\theta_0) dW(u) dW(v) \right]^2 \int k^4(z) dz} p[1 + o_P(1)].$$

Note that $\hat{q}$ goes to infinity at a rate $p$ as $p \to \infty$. For a very large $\hat{q}$, the asymptotic normality and the $\chi^2_{\hat{q}}$ approximation will deliver the same conclusion. When $\hat{q}$ is not large, there may be some difference in finite samples; we expect that the Chi-square approximation may perform better than normal approximation in finite sample with a moderate size of $\hat{q}$ because the Chi-square approximation may capture possible skewness of the finite sample distribution of $\hat{Q}$. We will investigate their finite sample performance via a simulation study.

We summarize the procedures to implement the tests $\hat{M}_1$ and $\hat{M}_2$ :

- Step 1: Obtain a $\sqrt{T}$-consistent estimator $\hat{\theta}$ for the model of interest.
- Step 2: Given a proper transformation, obtain and save the generalized residual $\hat{Z}_t = Z_t(\hat{\theta})$.
- Step 3: Compute the test statistic $\hat{M}_1$ in (3.11) or $\hat{M}_2$ in (3.12).
- Step 4: Compare $\hat{M}_1$ or $\hat{M}_2$ with an upper-tailed N(0,1) critical value (e.g., 1.65 at the 5% level), and reject $\mathbb{H}_0$ at a given level if $\hat{M}_1$ or $\hat{M}_2$ is larger than the critical value.

## 4. Asymptotic Theory

*4.1 Asymptotic Distribution*

We now investigate the asymptotic properties of the test statistics $\hat{M}_1$ and $\hat{M}_2$ under the null and alternative hypotheses. To derive the asymptotic distribution of $\hat{M}_1$ and $\hat{M}_2$ under $\mathbb{H}_0$, we impose the following regularity conditions.

**Assumption A.1:** (i) With probability one, $Z_t(\cdot) \equiv Z(I_t, \cdot)$ is twice continuously differentiable with respect to $\theta \in \Theta$ such that $\sup_{\theta\in\Theta} E||\frac{\partial}{\partial\theta} Z_t(\theta)||^2 \leq C$ and $\sup_{\theta\in\Theta} E||\frac{\partial^2}{\partial\theta\partial\theta'} Z_t(\theta)|| \leq C$; (ii) $\sup_{\theta\in\Theta} \frac{\partial}{\partial\theta} \|\varphi(u|\theta)\| < C$, where $\varphi(u|\theta) = E_\theta(e^{iuZ_t})$ and $E_\theta(\cdot)$ is the expectation operator under $F_\theta(\cdot)$; (iii) for each $\theta \in \Theta$, the process $\{Z_t(\theta), \frac{\partial}{\partial\theta'} Z_t(\theta)\}'$ is a strictly stationary $\alpha$-mixing process with the $\alpha$-mixing coefficient satisfying $\sum_{j=-\infty}^{\infty} \alpha(j)^{\frac{\nu-1}{\nu}} \leq C$ for some $\nu > 1$.

**Assumption A.2:** For a non-Markovian process $\{Y_t\}$, the information set $I_t = \{Y_t, Y_{t-1}, \cdots\}$ is infinite dimensional. Let $\hat{I}_t = \{Y_t, Y_{t-1}, \cdots, Y_1, \hat{J}_0\}$ be the observed information set available at period $t$ that may contain some assumed initial values $\hat{J}_0$. Then $\sup_{\theta\in\Theta} \sum_{t=1}^{T} \left| Z(\hat{I}_t, \theta) - Z(I_t, \theta) \right| = O_P(1)$.

**Assumption A.3:** $\sqrt{T}(\hat{\theta} - \theta_0) = O_P(1)$, where $\theta_0 \equiv p\lim(\hat{\theta}) \in \text{int}(\Theta)$, and $\theta_0$ is the same as in $\mathbb{H}_0$.

**Assumption A.4:** $k : \mathbb{R} \to [-1, 1]$ is symmetric continuous on all except a finite number of points in the real line. Furthermore, $k(0) = 1$ and $|k(z)| \leq C|z|^{-b}$ for $b > 3/2$ as $|z| \to \infty$.

**Assumption A.5:** $W : \mathbb{R} \to \mathbb{R}^+$ is a positive, nondecreasing, and right continuous weight function that weighs sets around zero equally, with $\int u^4 dW(u) \leq C$.

Assumption A.1(i) imposes regularity conditions on the generalized residual $Z_t(\theta)$ of the model $\mathcal{M}(\theta)$ for time series $\{Y_t\}$. Assumption A.1(ii) imposes a condition on the marginal distribution function $F_\theta(\cdot)$. Suppose $F_\theta(\cdot)$ is an absolutely continuous distribution with probability density function $f_\theta(z) = F'_\theta(z)$. Then Assumption A.1(ii) holds if $\sup E_\theta||\frac{\partial}{\partial\theta} \ln f_\theta(Z_t)|| \leq C$. Assumption A.1(iii) imposes some temporal dependence conditions on the related processes.

As pointed out earlier, we allow a non-Markovian process (e.g., GARCH models) for $\{Y_t\}$ and so the information set $I_t$ contains all its past history dating back to the infinite past. Since $I_t$ is infinite-dimensional, one may have to use a truncation version of $I_t$, that is, one has to use $\hat{I}_t = \{Y_t, Y_{t-1}, \cdots, Y_1, \hat{J}_0\}$, where $\hat{J}_0$ denotes some assumed initial values. For example, consider an AR(1)-GARCH(1,1)-*i.i.d.*N(0,1) model, which is a special case of (2.1) with $\mu_t = \alpha_0 + \alpha_1 Y_{t-1}$ and $h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2 (Y_{t-1} - \mu_{t-1})^2$. Then $\hat{J}_0 = (\hat{Y}_{-1}, \hat{Y}_0, \hat{h}_0)'$ are some assumed initial value for $Y_{-1}, Y_0$, and $h_0$ respectively. Assumption A.2 states that the truncation of the possibly infeasible information set $I_t$ has asymptotically negligible impact. Also see Bai (2003) for related discussion.

Assumption A.3 requires a $\sqrt{T}$-consistent estimator $\hat{\theta}$ under $\mathbb{H}_0$. We do not require any asymptotically most efficient estimator or a specified estimator. This is convenient for practitioners because for some nonlinear time series models, it is difficult to obtain asymptotically most efficient estimators. Assumption A.4 is the regularity condition on the kernel function. The continuity of $k(\cdot)$ at 0 and $k(0) = 1$ ensures that the bias of the generalized spectral estimator $\hat{f}(\omega, u, v)$ vanishes to zero asymptotically as $T \to \infty$. The condition on the tail behavior of $k(\cdot)$ ensures that higher order lags have asymptotically negligible impact on the statistical properties of $\hat{f}(\omega, u, v)$. Assumption A.4 covers most commonly used kernels (e.g., Priestley 1981, p.442). For kernels with bounded support, such as the Bartlett and Parzen kernels, $b = \infty$. For kernels with unbounded support, $b$ is some finite positive real number. For example, $b = 2$ for the Daniell and Quadratic-Spectral kernels. Finally, Assumption A.5 imposes mild conditions on the weighting function $W(\cdot)$. Any CDF with finite fourth moments satisfies Assumption A.5. In empirical characteristic function literature, it has been noted that any suitable positive, integrable and symmetric function will be sufficient for weight function (see, e.g., Huskova and Meintanis (2008)).

We now derive the asymptotic distribution of $\hat{M}_1$ and $\hat{M}_2$ under $\mathbb{H}_0$.

**Theorem 1:** *Suppose Assumptions A.1–A.5 hold, and $p = cT^{\lambda}$ for $\lambda \in (0, \frac{1}{2})$ and $c \in (0, \infty)$. Then under $\mathbb{H}_0$, $\hat{M}_1 - \hat{M}_2 \xrightarrow{p} 0$, $\hat{M}_1 \xrightarrow{d} N(0,1)$, and $\hat{M}_2 \xrightarrow{d} N(0,1)$ as $T \to \infty$.*

An important feature of $\hat{M}_1$ and $\hat{M}_2$ is their robustness to persistence in the original time series $\{Y_t\}$. This occurs because the generalized residual series $\{Z_t\}$ is always *i.i.d.* under $\mathbb{H}_0$ no matter how persistent serial dependence of $\{Y_t\}$ is. The robustness of the size performance avoids the use of bootstrap methods which may involve reestimating nonlinear time series models and are computationally costly. Moreover, some time series models such as Examples 6 and 7 in Section 2 do not fully specify the conditional distributions of $Y_t$. As a result, the usual parametric bootstrap cannot be used. Indeed, the robust inference based on asymptotic normality approximation is rather convenient in practice.

Intuitively, since a $\sqrt{T}$-consistent estimator $\hat{\theta}$ converges to $\theta_0$ under $\mathbb{H}_0$ faster than the nonparametric estimator $\hat{f}(\omega, u, v)$ converges to $f(\omega, u, v)$, the asymptotic distribution of $\hat{M}_1(p)$ is solely determined by the nonparametric estimator $\hat{f}(\omega, u, v)$. Consequently, the sampling variation of parameter estimator $\hat{\theta}$ has no impact on the asymptotic normal distribution. In other words, the asymptotic distribution of $\hat{M}_1$ remains unchanged when $\hat{\theta}$ is replaced by its probability limit $\theta_0$. This holds no matter whether the asymptotically most efficient estimator or a specific estimator is used. This asymptotic nuisance parameter free property leads to a convenient procedure. Our simulation study shows that the sampling error of $\hat{\theta}$ has little impact on the distribution of $\hat{M}_1$ and $\hat{M}_2$. This is particularly appealing for testing some nonlinear time series models because, for example, it has been well-known that efficient estimators of jump diffusion models and Markov chain regime-switching models are difficult to obtain.

Both $\hat{M}_1$ and $\hat{M}_2$ are asymptotically equivalent but $\hat{M}_2$ is a bit more convenient to compute. We will examine their finite sample performance via simulation.

### *4.2 Asymptotic Power*

We now investigate the asymptotic power property of the proposed tests under the alternatives to $\mathbb{H}_0$.

**Theorem 2:** *Suppose Assumptions A.1–A.5 hold, and* $p = cT^{\lambda}$ *for* $\lambda \in (0, \frac{1}{2})$ *and* $c \in (0, \infty)$*. Then under the alternative to* $\mathbb{H}_0$*, we have as* $T \to \infty$*, for* $i = 1, 2$*,*

$$\frac{\sqrt{p}}{T}\left[\int_{-\infty}^{\infty} k^2(z)dz\right]\hat{M}_i \xrightarrow{p} \iint_{-\pi}^{\pi} |f(\omega,u,v) - f_{\theta_0}(\omega,u,v)|^2 \, d\omega dW(u)dW(v)$$

$$= \frac{1}{2}\int |\sigma_0(u,v) - \sigma_0(u,v|\theta_0)|^2 dW(u)dW(v) + 2\sum_{j=1}^{\infty}\int |\sigma_j(u,v)|^2 dW(u)dW(v).$$

When $\{Z_t(\theta)\}$ is not *i.i.d.* or $Z_t(\theta)$ does not have the marginal distribution $F_\theta(z)$, $\hat{M}_1$ will have asymptotic power one (i.e., $\Pr(\hat{M}_1 > C) \to 1$ as $T \to \infty$ for any given constant $C$) provided that the weighting function $W(\cdot)$ is positive, monotonically nondecreasing and continuous with unbounded support on $\mathbb{R}$. Specifically, when the marginal distribution of $Z_t$ is not $F_\theta(z)$, the first term $\frac{1}{2}\int |\sigma_0(u,v) - \sigma_0(u,v|\theta_0)|^2 dW(u)dW(v)$ is positive. When $\{Z_t\}$ is not pairwise independent, the second term $2\sum_{j=1}^{\infty}\int |\sigma_j(u,v)|^2 \; dW(u)dW(v)$ is positive. Thus, we expect that $\hat{M}_1$ has relatively omnibus power against a wide variety of misspecification either in lag structure and parametric marginal distribution $F_\theta(z)$. This is confirmed in our simulation below.

*4.3 Data-Driven Bandwidth*

A practical issue in implementing our tests is the choice of lag order or bandwidth $p$. An advantage of our generalized spectral approach is that it can provide a data-driven method to choose $p$, which let data themselves determine a proper $p$ for $\hat{M}_1$ and $\hat{M}_2$. To justify the use of a data-driven lag order, $\hat{p}$ say, we impose a Lipschitz continuity condition on the kernel $k(\cdot)$. This condition rules out the truncated kernel $k(z) = \mathbf{1}(|z| \le 1)$, where $\mathbf{1}(\cdot)$ is the indicator function, but it still includes most commonly used kernels.

**Assumption A.6**: *For any* $x, y \in \mathbb{R}$, $|k(x) - k(y)| \le C|x - y|$ *for some constant* $C \in (0, \infty)$*.*

**Theorem 3**: *Suppose Assumptions A.1–A.6 hold, and* $\hat{p}$ *is a data-driven bandwidth such that* $\hat{p}/p = 1 + O_P(p^{-(\frac{3}{2}\beta - 1)})$ *for some* $\beta > (2b - \frac{1}{2})/(2b - 1)$*, where* $b$ *is as in Assumption A.4, and* $p$ *is a nonstochastic bandwidth with* $p = cT^{\lambda}$ *for* $\lambda \in (0, \frac{1}{2})$ *and* $c \in (0, \infty)$*. Then under* $\mathbb{H}_0$*,* $\hat{M}_i(\hat{p}) - \hat{M}_i(p) \xrightarrow{p} 0$ *and* $\hat{M}_i(\hat{p}) \xrightarrow{d} N(0, 1)$*,* $i = 1, 2$*.*

Theorem 3 implies that, as long as $\hat{p}$ converges to $p$ sufficiently fast, the use of $\hat{p}$ rather than $p$ has no impact on the limit distribution of $\hat{M}_1$. Theorem 3 allows for a wide range of admissible rates for $\hat{p}$. One plausible choice of $\hat{p}$ is the nonparametric plug-in method similar to that considered in Hong (1999). It is an estimation based optimal bandwidth (e.g, Härdle and Mammen (1993), Hjellvik and Tjøstheim (1995), Li (1999), Chen, Härdle and Li (2003)) in that it minimizes an asymptotic integrated mean square error (IMSE) criterion for the estimator $\hat{f}(\omega, u, v)$ (see, Hong (1999) Theorem 1). Nonparametric plug-in methods are not uncommon in the literature (e.g., Newey and West 1994, Silverman 1986). It considers some "pilot" generalized spectral derivative estimators based on a preliminary bandwidth $\bar{p}$ :

$$\bar{f}(\omega,u,v) \equiv \frac{1}{2\pi}\sum_{j=1-T}^{T-1}(1-|j|/T)^{\frac{1}{2}}\bar{k}(j/\bar{p})\hat{\sigma}_j(u,v)e^{-ij\omega}, \tag{4.1}$$

$$\bar{f}^{(q,0,0)}(\omega,u,v) \equiv \frac{1}{2\pi}\sum_{j=1-T}^{T-1}(1-|j|/T)^{\frac{1}{2}}\bar{k}(j/\bar{p})\hat{\sigma}_j(u,v)|j|^q e^{-ij\omega}, \tag{4.2}$$

where the kernel $\bar{k}:\mathbb{R}\to[-1,1]$ need not be the same as the kernel $k(\cdot)$ used in (3.8). For example, $\bar{k}(\cdot)$ can be the Bartlett kernel while $k(\cdot)$ is the Daniell kernel. Note that $\bar{f}(\omega,u,v)$ is an estimator for $f(\omega,u,v)$ and $\bar{f}^{(q,0,0)}(\omega,u,v)$ is an estimator for the generalized spectral derivative

$$f^{(q,0,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j(0,v)|j|^q e^{-ij\omega}. \tag{4.3}$$

Suppose for the kernel $k(\cdot)$, there exists some $q\in(0,\infty)$ such that $0<k^{(q)}\equiv\lim_{z\to 0}\frac{1-k(z)}{|z|^q}<\infty$. Then the plug-in bandwidth is defined as

$$\hat{p}_0 \equiv \hat{c}_0 T^{\frac{1}{2q+1}}, \tag{4.4}$$

where the tuning parameter estimator

$$\begin{aligned}\hat{c}_0 &\equiv \left[\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty}k^2(z)dz}\frac{\int\int_{-\pi}^{\pi}|\bar{f}^{(q,0,0)}(\omega,u,v)|^2 d\omega dW(u)dW(v)}{\int_{-\pi}^{\pi}[\int\bar{f}(\omega,v,-v)dW(v)]^2 d\omega}\right]^{\frac{1}{2q+1}}\\ &= \left[\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty}k^2(z)dz}\frac{\sum_{j=1-T}^{T-1}(T-|j|)\bar{k}^2(j/\bar{p})|j|^{2q}\int|\hat{\sigma}_j(u,v)|^2 dW(u)dW(v)}{\sum_{j=1-T}^{T-1}(T-|j|)\bar{k}^2(j/\bar{p})\operatorname{Re}\int\hat{\sigma}_j(u,-u)\hat{\sigma}_j(v,-v)dW(u)dW(v)}\right]^{\frac{1}{2q+1}}.\end{aligned}$$

The second equality here follows from Parseval's identity.

The data-driven $\hat{p}_0$ in (4.4) still involves the choice of a preliminary bandwidth $\bar{p}$, which can be either fixed or grow with the sample size $T$. If $\bar{p}$ is fixed, $\hat{p}_0$ still grows at rate $T^{\frac{1}{2q+1}}$ under $\mathbb{H}_A$ in general, but $\hat{c}_0$ does not converge to the optimal tuning constant that minimizes the IMSE of $\hat{f}(\omega,u,v)$. However, in practice, $\hat{c}_0$ will converge to some constant $\tilde{c}$ at the parametric rate. Thus, it is expected that $\hat{p}$ will easily satisfy the condition of Theorem 3. This is analogous in spirit to a parametric plug-in method. Following Hong (1999), we can show that when $\bar{p}$ grows with $T$ properly, the data-driven bandwidth $\hat{p}_0$ in (4.4) minimizes an asymptotic IMSE of $\hat{f}(\omega,u,v)$. Note that $\hat{p}_0$ is real-valued. One can take its integer part, and the impact of integer-clipping is expected to be negligible. The choice of $\bar{p}$ is somewhat arbitrary, but we expect that the choice of $\bar{p}$ is of secondary importance and may have no significant impact on $\hat{M}_a(\hat{p}_0)$. This is confirmed in our simulation below.

## 5. Monte Carlo Evidence

We now investigate the finite sample performance of the proposed tests.

*5.1 Simulation Design*

To examine the size performance of the tests $\hat{M}_1$ and $\hat{M}_2$ under $\mathbb{H}_0$, we consider the following data generating process (DGP):

DGP S.1 [AR(1)-GARCH(1,1)-$i.i.d.$N(0,1)]:

$$\begin{cases} Y_t = \alpha Y_{t-1} + u_t, \\ u_t = h_t^{1/2}\varepsilon_t, \\ \varepsilon_t \sim i.i.d.N(0,1), \end{cases}$$

where we set $\alpha = 0.2, 0.6, 0.9$ respectively to examine the impact of persistence in $\{Y_t\}$. For each $\alpha$, we consider a GARCH(1,1) specification:

$$h_t = 0.2 + \beta h_{t-1} + \gamma u_{t-1}^2,$$

with $(\beta, \gamma) = \{(0.6, 0.2), (0.79, 0.2), (0.8, 0.2)\}$ respectively. When $(\beta, \gamma) = (0.8, 0.2)$, $\{Y_t\}$ is an integrated GARCH process, which is strictly stationary but not weakly stationary.

We also considered the Chi-square approximations of $\hat{M}_1$ and $\hat{M}_2$, denoted $Q_1$ and $Q_2$ respectively (see discussion in Section 3). We expect that the Chi-square approximation may perform better than normal approximation in finite samples with a moderate size of degrees of freedom because the former may capture possible skewness of the finite sample distribution of the quadratic form $\hat{Q}$ in (3.10).

To examine the power of the tests, we consider the following DGPs:

DGP P.1 [AR(2)-GARCH(1,1)-$i.i.d.$N(0,1)]:

$$\begin{cases} Y_t = 0.2Y_{t-1} + 0.2Y_{t-2} + u_t, \\ u_t = h_t^{1/2}\varepsilon_t, \quad \{\varepsilon_t\} \sim i.i.d.N(0,1), \\ h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2, \end{cases}$$

DGP P.2 [TAR(1)-GARCH(1,1)-$i.i.d.$N(0,1)]:

$$\begin{cases} Y_t = -0.5Y_{t-1}\mathbf{1}(Y_{t-1} > 0) + 0.7Y_{t-1}\mathbf{1}(Y_{t-1} \leq 0) + u_t, \\ u_t = h_t^{1/2}\varepsilon_t, \quad \{\varepsilon_t\} \sim i.i.d.N(0,1), \\ h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2, \end{cases}$$

DGP P.3 [AR(1)-TGARCH(1,1)-$i.i.d.$N(0,1)]:

$$\begin{cases} Y_t = 0.2Y_{t-1} + u_t, \\ u_t = h_t^{1/2}\varepsilon_t, \quad \{\varepsilon_t\} \sim i.i.d.N(0,1), \\ h_t = 0.2 + 0.6h_{t-1} + 0.1u_{t-1}^2\mathbf{1}(u_{t-1} > 0) + 0.5u_{t-1}^2\mathbf{1}(u_{t-1} \leq 0). \end{cases}$$

DGP P.4 [AR(1)-GARCH(1,1)-$i.i.d.$ $EXP(1)$]:

$$\begin{cases} Y_t = 0.2Y_{t-1} + u_t, \\ u_t = h_t^{1/2}\varepsilon_t, \quad \{\varepsilon_t\} \sim i.i.d.\{\exp(1) - 1\}, \\ h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2. \end{cases}$$

DGP P.5 [AR(1)-GARCH(1,1)-$m.d.s.$ Innovations]:

$$\begin{cases} Y_t = 0.2Y_{t-1} + u_t, \\ u_t = h_t^{1/2}\varepsilon_t, \\ \varepsilon_t = \frac{\exp(\lambda_t\xi_t) - \exp(0.5\lambda_t^2)}{\sqrt{\exp(2\lambda_t^2) - \exp(\lambda_t^2)}}, \quad \{\xi_t\} \sim i.i.d.N(0,1), \\ \lambda_t^2 = 0.2 + 0.6\lambda_{t-1}^2 + 0.2u_{t-1}^2. \end{cases}$$

We generate data with the sample sizes $T = 250, 500$ and 1000 respectively. For each data set, we first generate $2T$ observations, and then discard the first $T$ ones to reduce the impact of some initial values. We then use MLE to estimate an AR(1)-GARCH(1,1)-$i.i.d.$ N(0,1) model:

$$\begin{cases} Y_t = \alpha Y_{t-1} + u_t, \\ u_t = h_t^{1/2}\varepsilon_t, \quad \{\varepsilon_t\} \sim i.i.d.N(0,1), \\ h_t = \phi + \beta h_{t-1} + \gamma u_{t-1}^2. \end{cases} \tag{5.1}$$

Under DGPs P.1 and P.2, model (5.1) suffers from dynamic misspecification in conditional mean. Under DGP P.3, which has a threshold effect in variance, model (5.1) suffers from a neglected nonlinearity in variance. DGP P.4 has a non-normal innovation distribution; it allows us to investigate misspecification in the marginal distribution of $\{\varepsilon_t\}$. We have also considered $\{\varepsilon_t\} \sim i.i.d.$ $\sqrt{\frac{3}{5}}t_{(5)}$ and mixed normal innovations ($\{\varepsilon_t\} \sim 0.5N(3,1) + 0.5N(-3,1)$). The results were similar to those of $\{\varepsilon_t\} \sim i.i.d.EXP(1)$. Under DGP P.5, $E(\varepsilon_t|I_{t-1}) = 0$ and $Var(\varepsilon_t|I_{t-1}) = 1$, but $\{\varepsilon_t\}$ is not $i.i.d.$ nor N(0,1). In particular, there exists serial dependence in conditional skewness and conditional kurtosis of $\{u_t\}$. We consider 10% and 5% significance levels. All results are obtained from 1000 iterations.

To compute $\hat{M}_1, \hat{M}_2, \hat{Q}_1$ and $\hat{Q}_2$ we use the N(0,1) CDF truncated on $[-3, 3]$ for the weighting function $W(\cdot)$, and use the Parzen kernel

$$k(z) = \begin{cases} 1 - 6z^2 + 6|z|^3 & \text{if } |z| \leq \frac{1}{2}, \\ 2(1 - |z|)^3 & \text{if } \frac{1}{2} \leq |z| < 1, \\ 0 & \text{otherwise,} \end{cases}$$

which has a bounded support and is computationally efficient. We have also used the Bartlett, Daniell and Quadratic Spectral kernels. The test statistics are similar to those based on the Parzen kernel in most cases. For the choice of lag order $p$, we use a data-driven lag order $\hat{p}_0$ via the plug-in method described in Hong (1999), with the Bartlett kernel $\bar{k}(z) = (1 - |z|)\mathbf{1}(|z| \leq 1)$ used in the preliminary generalized spectral derivative estimators. To certain extent, the data driven lag order $\hat{p}_0$ lets data tell an appropriate lag order, but it still involves the choice of the preliminary bandwidth $\bar{p}$ which is somewhat arbitrary. To examine the impact of the choice of the preliminary bandwidth $\bar{p}$, we consider $\bar{p} = 10, 15, 20, 25, 30$ respectively, which covers a sufficiently wide range of preliminary lag orders for the sample sizes $T$ considered here.

*5.2 Monte Carlo Evidence*

Tables 1-3 report the empirical rejection rates of the four tests under $\mathbb{H}_0$ at the 10% and 5% significance levels, using the asymptotic theory. Under DGP S.1 with a small autoregressive coefficient $\alpha = 0.2$, all tests $\hat{M}_1, \hat{M}_2, \hat{Q}_1$, and $\hat{Q}_2$ a bit underreject the null hypothesis $\mathbb{H}_0$, especially with low preliminary lag orders $\bar{p}$. But size improves as preliminary lag order $\bar{p}$ and sample size $T$ increase. The tests $\hat{M}_1$ and $\hat{Q}_1$ have better sizes than the tests $\hat{M}_2$ and $\hat{Q}_2$ in most cases, and the tests $\hat{M}_1$ and $\hat{M}_2$ have better sizes than their $\chi^2$ approximations, $\hat{Q}_1$ and $\hat{Q}_2$ respectively. Interestingly, with $\beta = 0.79$ and 0.8, high persistence in conditional variance, the tests $\hat{M}_1$ and $\hat{M}_2$ have similar sizes to those under $\beta = 0.6$. This is true even for an integrated GARCH(1,1) process ($\beta = 0.8$). This highlights the merits of the tests in that they are robust to the persistence in variance of the time series $\{Y_t\}$.

Under DPG S.1 with the medium autoregressive coefficient $\alpha = 0.6$, all tests have reasonable sizes with $T = 250$ and 500, regardless of persistence in the conditional variance. Overall, $\hat{M}_1$ and $\hat{Q}_1$ have better sizes than $\hat{M}_2$ and $\hat{Q}_2$, and $\hat{M}_1$ and $\hat{M}_2$ have similar sizes to $\hat{Q}_1$ and $\hat{Q}_2$. Under DGP S.1 with a large autoregressive coefficient $\alpha = 0.9$, $\hat{M}_1$ and $\hat{M}_2$ show some overrejection, especially with $\beta = 0.6$ and $\beta = 0.8$ at the 5% level. However, all tests have reasonable sizes with $T = 1000$. Now, the tests $\hat{Q}_1$ and $\hat{Q}_2$ have better sizes than $\hat{M}_1$ and $\hat{M}_2$ respectively, especially with larger sample sizes, and $\hat{M}_2$ has better sizes than $\hat{M}_1$.

We now compare the powers of the tests under DGP P.1-P.5, reported in Tables 4-6. Table 4 reports the empirical rejection rates of the tests at the 10% and 5% levels under DGPs P.1 and P.2 using the empirical critical values obtained under DGP S.1 with $\alpha = 0.2$ and $\beta = 0.6$, which provide a relatively fair ground to compare different tests. Under DGP P.1 (AR(2)), model (5.1) suffers from a linear dynamic misspecification in mean. All tests have similar powers at both significance levels and for all sample sizes. When $T = 250$, there is a slight tendency that power decreases as the preliminary lag order $\bar{p}$ increases. However, power becomes more robust to the level of $\bar{p}$ with larger sample sizes such as $T = 500$ and 1000. The tests have close to unit power for all tests with $T = 1000$. This confirms the merit of capturing dynamic misspecification of our test. Under DGP P.2 (TAR), there exists neglected nonlinearity in mean for model (5.1). All tests strongly reject the null hypothesis in all cases. With sample sizes $T = 500$ and 1000, all tests have essentially unit power.

Under DGP P.3 (TGARCH), model (5.1) suffers from neglected nonlinearity in conditional variance. The tests $\hat{M}_1$ and $\hat{Q}_1$ have similar powers to $\hat{M}_2$ and $\hat{Q}_2$ respectively, and $\hat{M}_1$ and $\hat{M}_2$ have slightly better powers than $\hat{Q}_1$ and $\hat{Q}_2$, but not substantially better. With a smaller sample size as $T = 250$, powers are not high, but they improve as sample size $T$ increases and become quite powerful when $T = 1000$. The powers of all tests decrease in $\bar{p}$. Under DGP P.4, we examine the impact of misspecification in the error distributions. Both $\hat{M}_1$ and $\hat{Q}_1$ have similar powers to $\hat{M}_2$ and $\hat{Q}_2$, respectively. The powers decrease in $\bar{p}$ substantially, especially with $T = 250$. This is consistent with our theory, since only the first term (with $j = 0$) in Theorem 2 captures the misspecification in the marginal distribution, and including more lags will result in a power loss when there exists no dynamic misspecification. However, with $T = 1000$, all tests have essentially unit power or close to it at both significance levels. This confirms that our test is powerful in capturing misspecification in the marginal error distribution.

Finally, under DGP P.5 (*m.d.s.* innovations), where there exists serial dependence in conditional skewness and conditional kurtosis of $\{\varepsilon_t\}$, all tests are very powerful and have similar power in all scenarios. Although there is some tendency that power decreases as $\bar{p}$ increases with $T = 250$, powers become more robust to the choice of $\bar{p}$ with larger sample sizes ($T = 500$ and $T = 1000$).

In summary, we have observed the following stylized facts:

- With low persistence in mean, the empirical sizes are smaller than the significance levels when the preliminary lag order $\bar{p}$ is small, but they improve as both the preliminary lag order $\bar{p}$ and the sample size $T$ increases. With medium and high persistence in mean, all tests have reasonable sizes.

- With high persistence in mean, the Chi-square approximated tests $\hat{Q}_1$ and $\hat{Q}_2$ perform better than $\hat{M}_1$ and $\hat{M}_2$.

- The sizes of all tests are relatively robust to the persistence in the conditional variance as well in the conditional mean.

- Our tests are powerful in detecting various model misspecifications, ranging from dynamic misspecification and neglected nonlinearity in mean and variance to marginal distribution misspecification and higher order dependence.

## 6. Conclusion

Using a generalized spectral approach, we have proposed a class of generally applicable new tests for nonlinear time series models based on the generalized residuals that are essentially a nonlinear filter transforming the original time series process into an *i.i.d.* sequence with a specified marginal distribution. This approach provides a unifying framework for testing various nonlinear time series models, including conditional probability distribution models, Markov-Chain regime switching models, conditional duration models, conditional intensity models, continuous-time jump diffusion models, continuous-time regression models, conditional quantile and interval models. The proposed test has a convenient asymptotic $N(0, 1)$ distribution which performs reasonably well in finite samples and is robust to dependent persistence in the original time series process. The test has relatively omnibus power against a wide range of model misspecifications via checking serial dependence over all lags and marginal distributions of the generalized residuals, as is illustrated by a simulation study.

# References


**Aït-Sahalia, Y.** (1996), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies* 9, 385-426.

**Bai, J.** (2003), "Testing Parametric Conditional Distributions of Dynamic Models," *Review of Economics and Statistics* 85, 531-549.

**Barndorff-Nielsen, O.E. and N. Shephard** (2001), "Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics," *Journal of the Royal Statistical Society, Series B* 63, 167-241.

**Box, G.E.P. and D.A. Pierce** (1970), "Distribution of the Residual Autocorrelations in ARIMA Time Series Models," *Journal of the American Statistical Association* 65, 1509-1526.

**Brillinger and Rosenblatt** (1967a), "Asymptotic Theory of k-th Order Spectra," in *Spectral Analysis of Time Series* (ed. B. Harris), New York : Wiley, 153-188.

**Brillinger and Rosenblatt** (1967b), "Computation and Interpretation of k-th Order Spectra" in *Spectral Analysis of Time Series* (ed. B. Harris), New York :Wiley, 189-232.

**Christoffersen, P.F.** (1998), "Evaluating Interval Forecasts", *International Economic Review*, 39, 841-862.

**Cox, D. R. and E.J. Snell** (1968), "A General Definition of Residuals," *Journal of the Royal Statistical Society* B 30, 248-275.

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

**Engle, R. F.** (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the U.K.Inflation," *Econometrica* 50, 987-1008.

**Engle, R.F. and J. Russell** (1998), "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data", *Econometrica* 66, p.1127-1162.

**Engle, R. F.** (2000), "The Econometrics of Ultra-High Frequency Data," *Econometrica* 68, 1-22.

**Engle, R. F. and Manganelli** (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles", *Journal of Business and Economic Statistic*s 22, 367-381.

**Fan, J. and Q.Yao** (2003), *Nonlinear Time Series*, Springer.

**Gao, J.** (2007), *Nonlinear Time Series*, Chapman and Hall.

**Granger, C.W.J. and T. Teräsvirta** (1993), *Modelling Nonlinear Economic Relationships.* Oxford University Press: New York.

**Hansen, B. E.** (1994), "Autoregressive Conditional Density Estimation," *International Economic Review* 35, 705-730.

**Hong, Y.** (1999), "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach," *Journal of the American Statistical Association* 84, 1201-1220.

**Hong, Y. and T.-H. Lee** (2003), "Diagnostic Checking for the Adequacy of Nonlinear Time Series Models," *Econometric Theory* 19, 1065-1121.

**Hong, Y. and H. Li** (2005), "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rate", *Review of Financial Studies* 18,

**Khmaladze, E.V.** (1981), "Martingale Approach in the theory of goodness-of-fit tests," *Theory of Probability and Its Applications* 26, 240–257.

**Koenker, R. and Z. Xiao** (2006), "Quantile Autoregression", *Journal of the American Statistical Association* 101, 980-990.

**Park, J. Y.** (2008), "Martingale Regression," Working Paper, Department of Economics, Texas A&M.

**Priestley, M.B.** (1981), *Spectral Analysis and Time Series.* Academic press: London.

**Rosenblatt, M.** (1952), "Remarks on a Multivariate Transformation", *Annals of Mathematical Statistics* 23, 470-472.

**Subba Rao, T. and M. Gabr** (1984), *An Introduction to Bispectral Analysis and Bilinear Time Series Models.* Lecture Notes in Statistics 24. Springer: New York.

**Thompson, S.** (2008), "Identifying Term Structure Volatility from the LIBOR-Swap Curve", *Review of Financial Studies* 21, 819-854.

**Tjøstheim, D.** (1994), "Non-Linear Time Series: A Selective Review," *Scandinavian Journal of Statistics* 21, 97-130.

**Tong, H.** (1990), *Nonlinear Time Series: A Dynamic System Approach*, Clarendon Press: Oxford.

**Yashin and Arjas** (1988), "A Note on Random Intensities and Conditional Survival Functions," *Journal of Applied Probability* 25, 630-635.

## MATHEMATICAL APPENDIX

Throughout the appendix, we let $M_1$ and $M_2$ be defined in the same way as $\hat{M}_1$ and $\hat{M}_2$ in (3.12)–(3.13), with the unobservable generalized residuals $\{Z_t \equiv Z_t(\theta_0)\}_{t=1}^T$, where $\theta_0 \equiv p\lim\hat{\theta}$, replacing the estimated generalized residuals $\{\hat{Z}_t \equiv Z(\hat{I}_{t-1},\hat{\theta})\}_{i=1}^T$. Also, $C \in (1,\infty)$ denotes a generic bounded constant.

**Proof of Theorem 1:** To show asymptotic equivalence between $\hat{M}_1$ and $\hat{M}_2$, it suffices to show that the difference of two centering factors of $\hat{M}_1$ and $\hat{M}_2$ is $o_P(1)$. The difference of two centering factors is $\int T^{-1}\sum_{t=1}^T|\hat{\varphi}_t(u)\hat{\varphi}_t(v)|^2$ $dW(u)dW(v)$, which is $O_P(1)$. Thus, $\hat{V}^{-1/2}\int T^{-1}\sum_{t=1}^T|\hat{\varphi}_t(u)\hat{\varphi}_t(v)|^2\,dW(u)dW(v) = o_P(1)$ as $p\to\infty$ given $\hat{V} = 2p[\iint\sigma_0(u,v|\theta_0)dW(u)dW(v)]^2\int_{-\infty}^{\infty}j^4(z)dz[1+o_P(1)] \propto p$.

To prove asymptotic normality of $\hat{M}_1$, it suffices to show Theorems A.1–A.2 below. Theorem A.1 implies that the use of $\{Z_t\}_{t=1}^T$ rather than $\{\hat{Z}_t\}_{t=1}^T$ has no impact on the limit distribution of $\hat{M}_1$.

Recall $\hat{\sigma}_j(u,v)$ as defined in (3.9) is the sample generalized autocovariance function of $\{\hat{Z}_t\}_{t=1}^T$. Let $\tilde{\sigma}_j(u,v)$ be defined in the same way as $\hat{\sigma}_j(u,v)$ with the unobservable generalized residuals $\{Z_t \equiv Z(I_{t-1},\theta_0)\}_{t=1}^T$. That is, define

$$\begin{aligned}\tilde{\sigma}_j(u,v) &= \frac{1}{T-|j|}\sum_{t=|j|+1}^{T}[e^{iuZ_t}-\tilde{\varphi}(u)][e^{ivZ_{t-|j|}}-\tilde{\varphi}(v)] \\ &= \frac{1}{T-|j|}\sum_{t=|j|+1}^{T}\tilde{\psi}_t(u)\tilde{\psi}_{t-|j|}(v), \end{aligned} \tag{A.1}$$

where $\tilde{\psi}_t(u) = e^{iuZ_t}-\tilde{\varphi}(u)$ and $\tilde{\varphi}(u) = T^{-1}\sum_{t=1}^T e^{iuZ_t}$.

**Theorem A.1:** *Under the conditions of Theorem 1,* $\hat{M}_1 - M_1 \xrightarrow{p} 0$.

**Theorem A.2:** *Under the conditions of Theorem 1,* $M_1 \xrightarrow{d} N(0,1)$.

**Proof of Theorem A.1:** To show that $\hat{M}_1 - M_1 \xrightarrow{p} 0$, it suffices to show that (i)

$$\begin{aligned}&\sum_{j=1-T}^{T-1}a_T(j)(T-|j|)\int\left|\hat{\sigma}_j(u,v)-\sigma_j(u,v|\hat{\theta})\right|^2 dW(u)dW(v) \\ =\ &\sum_{j=1-T}^{T-1}a_T(j)(T-|j|)\int\left|\tilde{\sigma}_j(u,v)-\sigma_j(u,v|\theta_0)\right|^2 dW(u)dW(v)+o_P(p^{1/2}),\end{aligned} \tag{A.2}$$

(ii) $\hat{A}-\tilde{A} = O_P(p/T^{1/2})$, and (iii) $p^{-1}(\hat{V}-\tilde{V}) \xrightarrow{p} 0$, where $\tilde{A}_1(p)$ and $\tilde{V}_1(p)$ are defined in the same way as $\hat{A}$ and $\hat{V}$ in (3.12), with $\{Z_t\}_{t=1}^T$ replacing $\{\hat{Z}_t\}_{t=1}^T$. For space, we focus on the proof of (A.2); the proofs for $\hat{A}-\tilde{A} = O_P(p/T^{\frac{1}{2}})$ and $p^{-1}(\hat{V}-\tilde{V}) \xrightarrow{p} 0$ are straightforward (though tedious). We note that it is necessary to obtain the convergence rate $O_P(p/T^{\frac{1}{2}})$ for $\hat{A}-\tilde{A}$ so as to ensure that replacing $\hat{A}$ with $\tilde{A}$ has asymptotically negligible impact given $p^2/T \to 0$.

Recall $\sigma_j(u,v|\theta) =\mathrm{cov}_\theta(e^{iuZ_t},e^{ivZ_{t-j}})$ is the generalized autocovariance function of $\{Z_t\}$ when $\{Z_t\}$ is *i.i.d.* $F_\theta(z)$. Thus, we have that for all $\theta\in\Theta$,

$$\sigma_j(u,v|\theta) = \begin{cases}\varphi(u+v|\theta)-\varphi(u|\theta)\varphi(v|\theta) & \text{if } j=0\\ 0 & \text{otherwise.}\end{cases}$$

Writing $\hat{\sigma}_j(u,v)-\sigma_j(u,v|\theta) = [\hat{\sigma}_j(u,v)-\tilde{\sigma}_j(u,v)]+[\tilde{\sigma}_j(u,v)-\sigma_j(u,v|\theta)]$, we can decompose

$$\left|\hat{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right|^2$$
$$= \left|\tilde{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right|^2 + |\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|^2 + 2\left[\tilde{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right][\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)]. \quad \text{(A.3)}$$

It follows that

$$\begin{aligned}
&\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int \left|\hat{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right|^2 dW(u)dW(v) \\
= &\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int \left|\tilde{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right|^2 dW(u)dW(v) \\
&+ \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int |\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|^2 dW(u)dW(v) \\
&+2\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int \left[\tilde{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right][\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)]\, dW(u).
\end{aligned}$$

We shall show the following propositions.

**Proposition A.1:** *Under the conditions of Theorem 1,*

$$\begin{aligned}
&\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int \left|\tilde{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right|^2 dW(u)dW(v) \\
= &\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int |\tilde{\sigma}_j(u,v) - \sigma_j(u,v|\theta_0)|^2 dW(u)dW(v) + O_p(1).
\end{aligned}$$

**Proposition A.2:** *Under the conditions of Theorem 1,*

$$\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int |\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|^2 dW(u)dW(v) = O_P(1).$$

**Proposition A.3:** *Under the conditions of Theorem 1,*

$$\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int \left[\tilde{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right][\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)]\, dW(u)dW(v) = o_P(p^{1/2}).$$

**Proof of Propostion A.1:** Noting that $\sigma_j(u,v|\theta) = 0$ for all $j \neq 0$ and all $\theta \in \Theta$, we have

$$\begin{aligned}
&\sum_{j=1-T}^{T-1} a_T(j)(T-j)\int \left|\tilde{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right|^2 dW(u)dW(v) \\
= &\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int |\tilde{\sigma}_j(u,v) - \sigma_j(u,v|\theta_0)|^2 dW(u)dW(v) \\
&+a_T(0)T\int \left|\sigma_0(u,v|\hat{\theta}) - \sigma_0(u,v|\theta_0)\right|^2 dW(u)dW(v) \\
&+2a_T(0)T\int [\tilde{\sigma}_0(u,v) - \sigma_0(u,v|\theta_0)]\left[\sigma_0(u,v|\hat{\theta}) - \sigma_0(u,v|\theta_0)\right] dW(u)dW(v)
\end{aligned}$$

$$= \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int |\tilde{\sigma}_j(u,v)-\sigma_j(u,v|\theta_0)|^2\, dW(u)dW(v) + O_P(1),$$

Here, we have used the fact that

$$\int \left|\sigma_0(u,v|\hat{\theta}) - \sigma_0(u,v|\theta_0)\right|^2 dW(u)dW(v) = O_P(T^{-1}) \tag{A.4}$$

by the mean value theorem, Assumptions A.1, A.3 and A.4. We also have used the fact that

$$\int [\tilde{\sigma}_0(u,v) - \sigma_0(u,v|\theta_0)] \left[\sigma_0(u,v|\hat{\theta}) - \sigma_0(u,v|\theta_0)\right] dW(u)dW(v) = O_P(T^{-1})$$

by the Cauchy-Swartz inequality, (A.4) and $\int |\tilde{\sigma}_0(u,v)-\sigma_0(u,v|\theta_0)|^2\, dW(u)dW(v) = O_P(T^{-1})$ under $\mathbb{H}_0$ by Markov's inequality. Note that we have $\sigma_0(u,v|\theta_0) = \sigma_0(u,v)$ under $\mathbb{H}_0$. ■

**Proof of Proposition A.2:** Put $\hat{\delta}_t(u) = e^{iu\hat{Z}_t} - e^{iuZ_t}$. Following the definitions of $\hat{\sigma}_j(u,v)$ and $\tilde{\sigma}_j(u,v)$, we decompose

$$\begin{aligned}
&(T-|j|)[\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)] \\
&= \sum_{t=j+1} \left[e^{iu\hat{Z}_t} - \hat{\varphi}(u)\right]\left[e^{iu\hat{Z}_{t-j}} - \hat{\varphi}(v)\right] - \tilde{\sigma}_j(u,v) \\
&= \sum_{t=|j|+1}^{T} \left[\left(e^{iu\hat{Z}_t} - e^{iuZ_t}\right) + \left(e^{iuZ_t} - \tilde{\varphi}(u)\right) + (\tilde{\varphi}(u) - \hat{\varphi}(u))\right] \\
&\quad \times \left[\left(e^{iv\hat{Z}_{t-|j|}} - e^{ivZ_{t-|j|}}\right) + \left(e^{ivZ_{t-|j|}} - \tilde{\varphi}(v)\right) + (\tilde{\varphi}(v) - \hat{\varphi}(v))\right] - \tilde{\sigma}_j(u,v) \\
&= \sum_{t=|j|+1}^{T} \hat{\delta}_t(u)\hat{\delta}_{t-|j|}(v) + \sum_{t=|j|+1}^{T} \hat{\delta}_t(u)\left(e^{ivZ_{t-|j|}} - \tilde{\varphi}(v)\right) + (\tilde{\varphi}(v) - \hat{\varphi}(v)) \sum_{t=|j|+1}^{T} \hat{\delta}_t(u) \\
&\quad + \sum_{t=|j|+1}^{T} \left(e^{iuZ_t} - \tilde{\varphi}(u)\right)\hat{\delta}_{t-|j|}(v) + (\tilde{\varphi}(v) - \hat{\varphi}(v)) \sum_{t=|j|+1} \left(e^{iuZ_t} - \tilde{\varphi}(u)\right) \\
&\quad + (\tilde{\varphi}(u) - \hat{\varphi}(u)) \sum_{t=|j|+1}^{T} \hat{\delta}_{t-|j|}(v) + (\tilde{\varphi}(u) - \hat{\varphi}(u)) \sum_{t=|j|+1}^{T} \left(e^{ivZ_{t-|j|}} - \tilde{\varphi}(v)\right) \\
&\quad + (\tilde{\varphi}(u) - \hat{\varphi}(u))(\tilde{\varphi}(v) - \hat{\varphi}(v)) \\
&= \sum_{c=1}^{8} \hat{B}_{cj}(u,v) \text{ say,}
\end{aligned} \tag{A.5}$$

We now show the order of magnitude of each term in (A.5). Lemmas A.1-A.8 are derived under the conditions of Theorem1.

**Lemma A.1:** $\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{1j}(u,v)\right|^2 dW(u)dW(v) = O_P(p/T)$.

**Lemma A.2:** $\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{2j}(u,v)\right|^2 dW(u)dW(v) = O_P(1)$.

**Lemma A.3:** $\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{3j}(u,v)\right|^2 dW(u)dW(v) = O_P(p/T)$.

**Lemma A.4:** $\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{4j}(u,v)\right|^2 dW(u)dW(v) = O_P(p/T)$.

**Lemma A.5:** $\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{5j}(u,v)\right|^2 dW(u)dW(v) = O_P(p^3/T^2)$.

**Lemma A.6:** $\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{6j}(u,v)\right|^2 dW(u)dW(v) = O_P(p/T)$.

**Lemma A.7:** $\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{7j}(u,v)\right|^2 dW(u)dW(v) = O_P(p^3/T^2)$.

**Lemma A.8:** $\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{8j}(u,v)\right|^2 dW(u)dW(v) = O_P(p/T)$.

**Proof of Lemma A.1:** For $\hat{B}_{1j}(u,v)$, by the inequality that $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ for any real $z_1$ and $z_2$, we have

$$\begin{aligned} \left|\sum_{t=|j|+1}^{T} (e^{iu\hat{Z}_t} - e^{iuZ_t})(e^{iv\hat{Z}_{t-|j|}} - e^{ivZ_{t-|j|}})\right| &\leq \sum_{t=|j|+1}^{T} |uv| \left|\hat{Z}_t - Z_t\right| \left|\hat{Z}_{t-|j|} - Z_{t-|j|}\right| \\ &\leq |uv| \sum_{t=1}^{T} (\hat{Z}_t - Z_t)^2. \end{aligned}$$

Note that

$$\begin{aligned} \sum_{t=1}^{T} (\hat{Z}_t - Z_t)^2 &\leq 2\sum_{t=1} \left[Z(\hat{I}_{t-1}, \hat{\theta}) - Z(I_{t-1}, \hat{\theta})\right]^2 + 2\sum_{t=1} \left[Z(I_{t-1}, \hat{\theta}) - Z(I_{t-1}, \theta_0)\right]^2 \\ &= O_P(1) + O_P(1) = O_P(1) \end{aligned}$$

where the first term is $O_P(1)$ by Assumption A.2, and the second term is $O_P(1)$ by the mean value theorem, Assumptions A.1 and A.3. It follows that

$$\begin{aligned} \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{1j}(u,v)\right|^2 dW(u)dW(v) &\leq \sum_{t=1}^{T} (\hat{Z}_t - Z_t)^2 \left[\int u^2 dW(u)\right]^2 \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \\ &= O_P(p/T), \end{aligned} \quad \text{(A.6)}$$

where, as shown in Hong (1999, (A15), p.1213), we have made use of the fact that

$$\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} = O_P(p/T). \blacksquare \quad \text{(A.7)}$$

**Proof of Lemma A.2:** Using the inequality that $|e^{iz} - 1 - iz| \leq |z|^2$ for any real $z$, we have

$$\left|e^{iu\hat{Z}_t} - e^{iuZ_t} - iu(\hat{Z}_t - Z_t)e^{iuZ_t}\right| \leq u^2(\hat{Z}_t - Z_t)^2. \quad \text{(A.8)}$$

By the second order Taylor series expansion, we have

$$\begin{aligned} \left|e^{iu\hat{Z}_t} - e^{iuZ_t} - iu\frac{\partial}{\partial\theta} Z_t(\theta_0) e^{iuZ_t}(\hat{\theta} - \theta_0)\right| &\leq u^2[\hat{Z}_t - Z_t(\theta_0)]^2 + |u||Z_t - Z_t(\hat{\theta})| \\ &\quad + |u|||\hat{\theta} - \theta_0||^2 \sup_{\theta\in\Theta} \left\|\frac{\partial^2}{\partial\theta\partial\theta'} Z_t(\theta)\right\|. \end{aligned} \quad \text{(A.9)}$$

Thus, given the fact that $|\tilde{\varphi}(v)| \leq C$, we obtain

$$\begin{aligned} (T-|j|)|\hat{B}_{2j}(u,v)| &\leq |u|||\hat{\theta} - \theta_0|| \left|\sum_{t=|j|+1}^{T} \tilde{\varphi}(v) \frac{\partial}{\partial\theta} Z_t(\theta_0) e^{iuZ_t}\right| + u^2 \sum_{t=1}^{T} [\hat{Z}_t - Z_t(\theta_0)]^2 \\ &\quad + |u| \sum_{t=1}^{T} [\hat{Z}_t - Z_t(\theta_0)]^2 + |u|||\hat{\theta} - \theta_0||^2 \sum_{t=1}^{T} \sup_{\theta\in\Theta} \left\|\frac{\partial^2}{\partial\theta\partial\theta'} Z_t(\theta)\right\|. \end{aligned} \quad \text{(A.10)}$$

It follows from (A.10) and Assumptions A.1-A.5 that

$$\begin{aligned}
&\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{2j}(u,v)\right|^2 dW(u)dW(v) \\
\leq\ & 8||\hat{\theta}-\theta_0||^2 \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left| \sum_{t=|j|+1}^{T} e^{iuZ_{t-|j|}} \frac{\partial}{\partial\theta} Z_t(\theta_0) e^{iuZ_t} \right|^2 u^2 dW(u)dW(v) \\
& +8\left[\sum_{t=1}^{T}\left(\hat{Z}_t - Z_t(\theta_0)\right)^2\right]^2 \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int u^4 dW(u)dW(v) \\
& +8\left[\sum_{t=1}^{T}\left(\hat{Z}_t - Z_t(\hat{\theta})\right)^2\right]^2 \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int u^2 dW(u)dW(v) \\
& +8||\hat{\theta}-\theta_0||^4 \left[\sum_{t=1}^{T} \sup_{\theta\in\Theta} \left\| \frac{\partial^2}{\partial\theta\partial\theta'} Z_t(\theta) \right\| \right]^2 \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int u^2 dW(u)dW(v) \\
=\ & O_P(1), \qquad \text{(A.11)}
\end{aligned}$$

where the last three terms are $O_P(p/T)$ given Assumptions A.1-A.5, and the first term is $O_P(1)$ following analogous reasoning to Hong and Lee (2003, Lemma A.5), based on the mixing property of $\{Z_t(\theta), \frac{\partial}{\partial\theta} Z_t(\theta)\}'$. ■

**Proof of Lemma A.3:** By the inequality that $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ for any real $z_1$ and $z_2$, we have

$$|\hat{\varphi}(v) - \tilde{\varphi}(v)| = \left| T^{-1} \sum_{t=1}^{T} (e^{iv\hat{Z}_t} - e^{ivZ_t}) \right| \leq |v| T^{-1} \sum_{t=1}^{T} \left| \hat{Z}_t - Z_t \right|,$$

and $\left|\sum_{t=|j|+1} \hat{\delta}_{t-j}(u)\right| \leq |u| \sum_{t=1} |\hat{Z}_t - Z_t|$. On the other hand,

$$\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}\left|\hat{Z}_t - Z_t\right| &\leq T^{-1/2}\sum_{t=1}^{T}\left|Z(\hat{I}_{t-1},\hat{\theta}) - Z(I_{t-1},\hat{\theta})\right| + T^{-1/2}\sum_{t=1}^{T}\left|Z(I_{t-1},\hat{\theta}) - Z(I_{t-1},\theta_0)\right| \\
&= O_P(T^{-1/2}) + O_P(1) = O_P(1)
\end{aligned}$$

where the first term is $O_P(T^{-1/2})$ by Assumption A.2, and the second term is $O_P(1)$ by the mean value theorem, Assumptions A.1 and A.3. It follows that

$$\begin{aligned}
\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{3j}(u,v)\right|^2 dW(u)dW(v) &\leq \left[T^{-1/2}\sum_{t=1}^{T}\left|\hat{Z}_t - Z_t\right|\right]^4 \left[\int u^2 dW(u)\right]^2 \sum_{j=1}^{T-1} a_T(j)(T-j)^{-1} \\
&= O_P(p/T). \ ■ \qquad \text{(A.12)}
\end{aligned}$$

**Proof of Lemma A.4:** By the similar reasoning to the proof of Lemma A.2, using (A.10), we have

$$\begin{aligned}
&\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{4j}(u,v)\right|^2 dW(u)dW(v) \\
\leq\ & 8||\hat{\theta}-\theta_0||^2 \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left| \sum_{t=|j|+1}^{T} e^{iuZ_t} \frac{\partial}{\partial\theta} Z_{t-|j|}(\theta_0) e^{iuZ_{t-|j|}} \right|^2 u^2 dW(u)dW(v) \\
& +8\left[\sum_{t=|j|+1}^{T} \left(\hat{Z}_{t-|j|} - Z_{t-|j|}(\theta_0)\right)^2\right]^2 \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int u^4 dW(u)dW(v) \\
& +8\left[\sum_{t=|j|+1}^{T} \left(\hat{Z}_{t-|j|} - Z_{t-|j|}(\hat{\theta})\right)^2\right]^2 \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int u^2 dW(u)dW(v) \\
& +8||\hat{\theta}-\theta_0||^4 \left[\sum_{t=|j|+1}^{T} \sup_{\theta\in\Theta} \left\| \frac{\partial^2}{\partial\theta\partial\theta'} Z_{t-|j|}(\theta) \right\| \right]^2 \sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int u^2 dW(u)dW(v) \\
=\ & O_P(p/T), \qquad \text{(A.13)}
\end{aligned}$$

where we have used the fact that $E\left|\sum_{t=|j|+1}^{T} e^{iuZ_t} \frac{\partial}{\partial\theta} Z_{t-|j|}(\theta) e^{iuZ_{t-|j|}}\right|^2 \leq C(T-|j|)$ because $e^{iuZ_t}$ is independent of $\frac{\partial}{\partial\theta} Z_{t-|j|}(\theta) e^{iuZ_{t-|j|}}$ for $j>0$ under the *i.i.d.* property of $\{Z_t\}_{t=1}^T$ under $\mathbb{H}_0$. ■

**Proof of Lemma A.5:** Since $\tilde{\varphi}(v) = T^{-1}\sum_{t=1}^{T} e^{iuZ_t}$, we have

$$\begin{aligned}
\left| [\hat{\varphi}(v) - \tilde{\varphi}(v)] \sum_{t=j+1}^{T} \left[e^{iuZ_t} - \tilde{\varphi}(u)\right] \right| &\leq |\hat{\varphi}(v) - \tilde{\varphi}(v)| \left| \sum_{t=1}^{j} \left[e^{iuZ_t} - \tilde{\varphi}(u)\right] \right| \\
&\leq |\hat{\varphi}(v) - \tilde{\varphi}(v)| \left| \sum_{t=1}^{j} \left[e^{iuZ_t} - \tilde{\varphi}(u)\right] \right| \leq 2\left|\hat{\varphi}(v) - \tilde{\varphi}(v)\right| \cdot j.
\end{aligned}$$

It follows that

$$\begin{aligned}
\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{5j}(u,v)\right|^2 dW(u)dW(v) &= \sum_{j=1-T}^{T-1} a_T(j) j^2 (T-|j|)^{-1} \left|\hat{\varphi}(v) - \tilde{\varphi}(v)\right|^2 \\
&= O_P(p^3/T^2), \qquad \text{(A.14)}
\end{aligned}$$

by using the fact that $p^{-1}\sum_{j=1}^{T-1} z^2 k^2(z) \to \int_0^\infty z^2 k^2(z) dz$ given Assumption A.4. ■

**Proof of Lemma A.6:** By the similar reasoning to Lemma A.2, we have

$$\left| \frac{1}{T} \sum_{t=|j|+1}^{T} [\hat{\varphi}(u) - \tilde{\varphi}(u)] \left(e^{iv\hat{Z}_{t-|j|}} - e^{ivZ_{t-|j|}}\right) \right| \leq |\hat{\varphi}(u) - \tilde{\varphi}(u)|\,|\hat{\varphi}(v) - \tilde{\varphi}(v)| = |uv| O_P(T^{-1}).$$

It follows that

$$\begin{aligned}
\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1} \int \left|\hat{B}_{6j}(u,v)\right|^2 dW(u)dW(v) &\leq \left(T^{-1/2}\sum_{t=1}^{T} \left|\hat{Z}_t - Z_t\right|\right)^2 \left[\int u^2 dW(u)\right]^2 \sum_{j=1}^{T-1} a_T(j)(T-j)^{-1} \\
&= O_P(p/T). \ ■ \qquad \text{(A.15)}
\end{aligned}$$

**Proof of Lemma A.7:** By the similar reasoning as Lemma A.5, we have

$$\left|[\hat{\varphi}(u)-\tilde{\varphi}(u)]\sum_{t=|j|+1}^{T}\left(e^{ivZ_{t-|j|}}-\tilde{\varphi}(v)\right)\right| \le |\hat{\varphi}(u)-\tilde{\varphi}(u)|\left|\sum_{t=1}^{j}[e^{ivZ_{t-|j|}}-\tilde{\varphi}(u)]\right| \le 2\,|\hat{\varphi}(u)-\tilde{\varphi}(u)|\cdot j.$$

It follows that

$$\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1}\int\left|\hat{B}_{7j}(u,v)\right|^2 dW(u)dW(v) = O_P(p^3/T^2). \blacksquare \tag{A.16}$$

**Proof of Lemma A.8:** By the similar reasoning to Lemma A.2, we have

$$\left|\frac{1}{T}\sum_{t=|j|+1}^{T}[\hat{\varphi}(u)-\tilde{\varphi}(u)]\,[\hat{\varphi}(v)-\tilde{\varphi}(v)]\right| = |[\hat{\varphi}(u)-\tilde{\varphi}(u)]\,[\hat{\varphi}(v)-\tilde{\varphi}(v)]| = |uv|O_P(T^{-1}).$$

It follows that

$$\begin{aligned}\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)^{-1}\int\left|\hat{B}_{8j}(u,v)\right|^2 dW(u)dW(v) &\le \left(T^{-1}\sum_{t=1}^{T}\left|\hat{Z}_t-Z_t\right|\right)^2\left[\int u^2 dW(u)\right]^2\sum_{j=1}^{T-1}a_T(j)(T-|j|)\\ &= O_P(p/T). \blacksquare \end{aligned}\tag{A.17}$$

Collecting (A.5)-(A.17), we obtain the desired result. ■

**Proof of Proposition A.3:**

$$\begin{aligned}&\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int\left[\tilde{\sigma}_j(u,v)-\sigma_j(u,v|\hat{\theta})\right][\hat{\sigma}_j(u,v)-\tilde{\sigma}_j(u,v)]\,dW(u)dW(v)\\ =\;& a_T(0)T\int\left[\tilde{\sigma}_0(u,v)-\sigma_0(u,v|\hat{\theta})\right][\hat{\sigma}_0(u,v)-\tilde{\sigma}_0(u,v)]\,dW(u)dW(v)\\ &+2\sum_{j=1}^{T-1}a_T(j)(T-j)\int\tilde{\sigma}_j(u,v)\,[\hat{\sigma}_j(u,v)-\tilde{\sigma}_j(u,v)]\,dW(u)dW(v)\\ =\;& O_P(1)+o_P(p^{1/2}).\end{aligned}$$

Here, the first term is $O_P(1)$ by the Cauchy-Scwartz inequality, $\int\left|\tilde{\sigma}_0(u,v)-\sigma_0(u,v|\hat{\theta})\right|^2 dW(u)dW(v) = O_P(T^{-1})$ and $\int|\hat{\sigma}_0(u,v)-\tilde{\sigma}_0(u,v)|^2 dW(u)dW(v) = O_P(T^{-1})$. The second term is $o_P(p^{1/2})$ by following analogous reasoning of Hong and Lee (2003, Prop. A.2, p.1111). ■

**Proof of Theorem A.2:** Following the proof of Hong and Lee (2003, Theorem A.2), we have

$$\left[\sum_{j=1-T}^{T-1} a_T(j)(T-|j|)\int|\hat{\sigma}_j(u,v)-\sigma_j(u,v)|^2 dW(u)dW(v)-\hat{A}\right]/\sqrt{\hat{V}} \xrightarrow{d} N(0,1),$$

where $\hat{A}$ and $\hat{V}$ are given in (3.12). ■

**Proof of Theorem 2**: The proof of Theorem 2 consists of the proofs of Theorems A.3 and A.4 below.

**Theorem A.3**: *Under the conditions of Theorem 2,* $(p^{\frac{1}{2}}/T)[\hat{M}_1-M_1] \xrightarrow{p} 0$.

**Theorem A.4:** *Under the conditions of Theorem 2,*

$$
\begin{aligned}
& (p^{\frac{1}{2}}/T)M(p) \xrightarrow{p} \iint_{-\pi}^{\pi} |f(\omega,u,v) - f_0(\omega,u,v)|^2 \, d\omega dW(u)dW(v) \\
= & \ \frac{1}{2}\int |\sigma_0(u,v) - \sigma_0(u,v|\theta_0)|^2 dW(u)dW(v) + 2\sum_{j=1}^{\infty}\int |\sigma_j(u,v)|^2 dW(u)dW(v).
\end{aligned}
$$

**Proof of Theorem A.3:** It suffices to show that

$$
\int \sum_{j=1}^{T-1} a_T(j)(T-|j|)\left|\hat{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right|^2 dW(u)dW(v) \xrightarrow{p} 0, \tag{A.18}
$$

$T^{-1}(\hat{A}-\tilde{A}) \xrightarrow{p} 0$, and $p^{-1}(\hat{V}-\tilde{V}) \xrightarrow{p} 0$, where $\tilde{A}$ and $\tilde{V}$ are defined in the same way as $\hat{A}$ and $\hat{V}$ in (3.12), with $\{Z_t\}_{t=1}^{T}$ replacing $\{\hat{Z}_t\}_{t=1}^{T}$. Since the proofs for $T^{-1}(\hat{A}-\tilde{A}) \xrightarrow{p} 0$, and $p^{-1}(\hat{V}-\tilde{V}) \xrightarrow{p} 0$ are straightforward, we focus on the proof of (A.18). Decompose

$$
\hat{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta}) = \sigma_j(u,v) - \sigma_j(u,v|\theta_0) + [\hat{\sigma}_j(u,v) - \sigma_j(u,v)] - [\sigma_j(u,v|\hat{\theta}) - \sigma_j(u,v|\theta_0)]
$$

We have

$$
\begin{aligned}
& \left|\hat{\sigma}_j(u,v) - \sigma_j(u,v|\hat{\theta})\right|^2 \\
= & \ |\sigma_j(u,v) - \sigma_j(u,v|\theta_0)|^2 + \left|[\hat{\sigma}_j(u,v) - \sigma_j(u,v)] - [\sigma_j(u,v|\hat{\theta}) - \sigma_j(u,v|\theta_0)]\right|^2 \\
& +2\left[\sigma_j(u,v) - \sigma_j(u,v|\theta_0)\right]\left\{[\hat{\sigma}_j(u,v) - \sigma_j(u,v)] - [\sigma_j(u,v|\hat{\theta}) - \sigma_j(u,v|\theta_0)]\right\}.
\end{aligned} \tag{A.19}
$$

By the Cauchy-Schwarz inequality, it suffices to show that

$$
\sum_{j=1}^{T-1} a_T(j)(T-|j|)\int \left|[\hat{\sigma}_j(u,v) - \sigma_j(u,v)] - [\sigma_j(u,v|\hat{\theta}) - \sigma_j(u,v|\theta_0)]\right|^2 dW(u)dW(v) = o_P(1).
$$

We decompose

$$
\begin{aligned}
& \sum_{j=1}^{T-1} a_T(j)(T-|j|)\int \left|[\hat{\sigma}_j(u,v) - \sigma_j(u,v)] - [\sigma_j(u,v|\hat{\theta}) - \sigma_j(u,v|\theta_0)]\right|^2 dW(u)dW(v) \\
\leq & \ 2\sum_{j=1}^{T-1} a_T(j)(T-|j|)\int |\hat{\sigma}_j(u,v) - \sigma_j(u,v)|^2 \, dW(u)dW(v) \\
& +2\sum_{j=1}^{T-1} a_T(j)(T-|j|)\int \left|\sigma_j(u,v|\hat{\theta}) - \sigma_j(u,v|\theta_0)\right|^2 dW(u)dW(v).
\end{aligned}
$$

For the first term, we have

$$
\sum_{j=1}^{T-1} a_T(j)(T-|j|)\int |\hat{\sigma}_j(u,v) - \sigma_j(u,v)|^2 \, dW(u)dW(v) = O(T^{-1}),
$$

following the proof similar to Hong and Lee (2003, Theorem A.3). For the second term, we have

$$
\begin{aligned}
& \sum_{j=1}^{T-1} a_T(j)(T-|j|)\int \left|\sigma_j(u,v|\hat{\theta}) - \sigma_j(u,v|\theta_0)\right|^2 dW(u)dW(v) \\
= & \ \sum_{j=1}^{T-1} a_T(j)(T-|j|)\int \left\|\frac{\partial}{\partial\theta}\varphi(u|\theta)\right\|^2 ||\hat{\theta}-\theta_0||^2 dW(u)dW(v) = O(T^{-1}),
\end{aligned}
$$

by the mean-value theorem. This completes the proof for Theorem A.3. ■

**Proof of Theorem A.4:** The proof is very similar to Hong (1999, Proof of Theorem 5), for the case $(m, l) = (0, 0)$. ■

## TABLE 1. Empirical Size of Tests

DGP S.1: $Y_t = 0.2Y_{t-1} + u_t$, $\varepsilon_t = h_t^{1/2}\varepsilon_t$

| | A: $h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2$ | | | | | | | | B: $h_t = 0.2 + 0.8h_{t-1} + 0.2u_{t-1}^2$ | | | | | | | | C: $h_t = 0.2 + 0.79h_{t-1} + 0.2u_{t-1}^2$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | |
| $p$ | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| $T = 250$ | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | 6.9 | 4.6 | 6.5 | 4.0 | 5.7 | 4.0 | 5.5 | 3.1 | 8.0 | 5.2 | 7.7 | 4.3 | 6.7 | 4.3 | 6.6 | 3.3 | 8.1 | 5.1 | 7.4 | 4.0 | 6.9 | 4.0 | 6.6 | 3.4 |
| 15 | 7.9 | 5.1 | 7.7 | 4.6 | 7.0 | 4.6 | 6.8 | 4.1 | 8.7 | 5.5 | 8.4 | 4.8 | 7.5 | 5.0 | 7.3 | 4.1 | 8.8 | 5.7 | 8.2 | 4.8 | 7.5 | 4.9 | 7.3 | 4.2 |
| 20 | 8.3 | 5.8 | 7.9 | 5.4 | 7.6 | 5.4 | 6.9 | 5.1 | 9.0 | 5.9 | 8.5 | 5.6 | 7.7 | 5.6 | 7.3 | 4.4 | 8.9 | 6.0 | 8.3 | 5.6 | 7.8 | 5.7 | 7.3 | 4.6 |
| 25 | 9.1 | 6.1 | 8.7 | 5.4 | 8.0 | 5.4 | 7.7 | 5.0 | 9.7 | 6.1 | 9.2 | 5.6 | 8.2 | 5.6 | 8.0 | 5.2 | 9.6 | 6.2 | 9.1 | 5.8 | 8.5 | 5.8 | 8.0 | 4.8 |
| 30 | 9.3 | 6.7 | 9.1 | 5.9 | 8.6 | 6.2 | 8.1 | 5.6 | 9.5 | 7.0 | 9.2 | 6.2 | 8.4 | 6.3 | 8.2 | 5.6 | 9.5 | 6.7 | 9.2 | 6.1 | 8.6 | 6.1 | 8.1 | 5.5 |
| $T = 500$ | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | 7.6 | 5.6 | 7.3 | 5.0 | 6.4 | 5.0 | 6.1 | 4.1 | 7.5 | 5.6 | 7.3 | 4.8 | 6.9 | 4.8 | 6.5 | 4.3 | 7.5 | 5.5 | 7.3 | 5.0 | 6.8 | 5.2 | 6.5 | 4.5 |
| 15 | 8.3 | 5.7 | 7.9 | 5.0 | 7.1 | 5.3 | 6.9 | 4.9 | 7.8 | 6.0 | 7.6 | 5.3 | 6.9 | 5.3 | 6.7 | 4.8 | 7.7 | 6.1 | 7.4 | 5.3 | 7.0 | 5.4 | 6.9 | 4.9 |
| 20 | 8.8 | 6.1 | 8.5 | 5.7 | 7.9 | 5.7 | 7.7 | 4.7 | 8.3 | 6.0 | 8.1 | 5.3 | 7.7 | 5.6 | 7.5 | 5.0 | 8.5 | 6.2 | 8.3 | 5.6 | 7.7 | 5.7 | 7.6 | 5.1 |
| 25 | 8.8 | 6.2 | 8.7 | 5.9 | 8.4 | 6.0 | 7.7 | 5.6 | 8.7 | 6.4 | 8.6 | 5.8 | 8.0 | 5.8 | 7.9 | 5.4 | 8.9 | 6.6 | 8.4 | 5.8 | 8.0 | 5.9 | 7.9 | 5.4 |
| 30 | 9.3 | 6.3 | 9.0 | 5.8 | 8.2 | 5.8 | 8.1 | 5.3 | 9.5 | 6.4 | 9.1 | 6.0 | 8.6 | 6.0 | 8.3 | 5.3 | 9.4 | 6.6 | 9.2 | 6.0 | 8.5 | 6.0 | 8.3 | 5.5 |
| $T = 1000$ | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | 6.3 | 4.4 | 5.8 | 4.5 | 5.2 | 3.5 | 4.7 | 2.9 | 6.7 | 4.1 | 6.3 | 3.5 | 5.6 | 3.5 | 5.4 | 3.2 | 6.6 | 4.0 | 6.1 | 3.5 | 5.7 | 3.5 | 5.4 | 3.2 |
| 15 | 6.6 | 4.4 | 6.1 | 3.7 | 5.7 | 3.7 | 5.5 | 3.0 | 6.9 | 4.3 | 6.8 | 4.0 | 6.4 | 4.0 | 6.1 | 3.4 | 7.1 | 4.4 | 6.7 | 3.9 | 6.3 | 3.9 | 6.1 | 3.4 |
| 20 | 6.8 | 4.5 | 6.6 | 4.0 | 6.1 | 4.0 | 5.7 | 3.6 | 7.5 | 4.5 | 7.1 | 4.3 | 6.4 | 4.3 | 6.0 | 3.6 | 7.5 | 4.4 | 7.2 | 4.3 | 6.4 | 4.3 | 5.9 | 3.8 |
| 25 | 7.5 | 4.9 | 7.3 | 4.6 | 6.6 | 4.7 | 6.3 | 3.6 | 7.9 | 5.2 | 7.4 | 4.4 | 7.0 | 4.4 | 6.6 | 4.0 | 7.9 | 5.2 | 7.6 | 4.3 | 7.2 | 4.5 | 6.5 | 4.0 |
| 30 | 7.0 | 4.9 | 7.4 | 4.8 | 6.9 | 4.8 | 6.5 | 4.4 | 8.4 | 5.3 | 7.9 | 5.0 | 7.4 | 5.0 | 7.2 | 4.4 | 8.3 | 5.5 | 8.1 | 4.9 | 7.4 | 5.0 | 7.0 | 4.3 |

Notes : (i) 1000 iterations;
(ii) $\hat{M}_1, \hat{M}_2$, generalized spectral tests, $Q_1$ and $Q_2$ are their Chi-square approximation, respectively;
(iii) The Parzen kernel is used for both $\hat{M}_1$ and $\hat{M}_2$; $p = 10, 15, 20, 25, 30$.
(iv) A: $Y_t = 0.2Y_{t-1} + u_t$, $u_t = h_t^{1/2}\varepsilon_t, h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2$; B: $h_t = 0.2 + 0.8h_{t-1} + 0.2u_{t-1}^2$;
C: $h_t = 0.2 + 0.79h_{t-1} + 0.2u_{t-1}^2$, $\varepsilon_t \sim i.i.d.$ N(0, 1).

**TABLE 2. Empirical Sizes of Tests (cont.)**

DGP S.1: $Y_t = 0.6Y_{t-1} + u_t$, $u_t = h_t^{1/2}\varepsilon_t$

| | A: $h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2$ | | | | | | | | B: $h_t = 0.2 + 0.8h_{t-1} + 0.2u_{t-1}^2$ | | | | | | | | C: $h_t = 0.2 + 0.79h_{t-1} + 0.2u_{t-1}^2$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | |
| $p$ | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| $T = 250$ | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | 10.7 | 6.8 | 10.2 | 5.2 | 8.7 | 5.2 | 8.1 | 4.1 | 10.1 | 6.4 | 9.6 | 4.9 | 8.9 | 5.0 | 8.4 | 4.4 | 10.6 | 6.3 | 9.9 | 5.0 | 8.8 | 5.0 | 8.5 | 4.5 |
| 15 | 10.8 | 7.1 | 10.6 | 5.8 | 9.3 | 5.9 | 8.7 | 4.9 | 10.2 | 7.4 | 10.1 | 6.5 | 9.1 | 6.5 | 9.0 | 4.8 | 10.2 | 7.3 | 9.8 | 6.4 | 9.0 | 6.4 | 8.8 | 4.8 |
| 20 | 10.5 | 7.2 | 10.1 | 6.0 | 9.1 | 6.1 | 8.7 | 5.7 | 10.4 | 7.4 | 10.1 | 6.5 | 9.2 | 6.7 | 9.1 | 5.4 | 10.3 | 7.4 | 10.0 | 6.2 | 9.2 | 6.4 | 8.8 | 5.4 |
| 25 | 10.9 | 6.9 | 10.4 | 6.2 | 9.8 | 6.3 | 9.3 | 5.8 | 11.1 | 7.3 | 10.5 | 6.2 | 9.4 | 6.3 | 9.3 | 5.7 | 10.8 | 7.2 | 10.0 | 6.1 | 9.5 | 6.1 | 9.2 | 5.7 |
| 30 | 11.5 | 7.2 | 10.9 | 6.3 | 10.0 | 6.4 | 9.9 | 5.9 | 11.3 | 7.4 | 11.0 | 6.0 | 10.1 | 6.0 | 9.7 | 5.7 | 11.4 | 7.2 | 11.2 | 6.0 | 10.0 | 6.3 | 9.7 | 5.7 |
| $T = 500$ | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | 8.5 | 6.8 | 8.1 | 6.3 | 7.9 | 6.3 | 7.8 | 4.9 | 8.4 | 6.5 | 8.2 | 5.8 | 8.0 | 5.9 | 7.6 | 5.1 | 8.4 | 6.6 | 8.3 | 5.8 | 7.4 | 5.4 | 7.6 | 5.1 |
| 15 | 8.7 | 7.0 | 8.6 | 6.5 | 8.1 | 6.5 | 8.0 | 6.4 | 9.0 | 7.0 | 8.6 | 6.7 | 8.2 | 6.7 | 8.1 | 5.8 | 8.9 | 7.0 | 8.7 | 6.5 | 8.1 | 6.2 | 8.0 | 6.1 |
| 20 | 9.6 | 6.8 | 9.4 | 6.6 | 8.5 | 6.6 | 8.0 | 6.4 | 10.2 | 7.0 | 10.0 | 6.5 | 8.8 | 6.7 | 8.4 | 5.9 | 10.0 | 7.0 | 9.7 | 6.5 | 8.3 | 6.6 | 8.3 | 6.2 |
| 25 | 10.2 | 7.0 | 9.6 | 6.7 | 9.3 | 6.7 | 8.9 | 6.1 | 10.8 | 7.3 | 10.6 | 6.3 | 9.9 | 6.6 | 9.3 | 5.5 | 10.8 | 7.2 | 10.2 | 6.5 | 9.3 | 6.7 | 9.3 | 5.8 |
| 30 | 10.8 | 7.3 | 10.3 | 6.5 | 9.6 | 6.5 | 9.3 | 6.1 | 11.5 | 7.6 | 10.9 | 6.6 | 10.2 | 6.7 | 10.0 | 5.8 | 11.2 | 7.8 | 10.7 | 6.9 | 9.7 | 6.6 | 9.8 | 5.9 |
| $T = 1000$ | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | 6.6 | 5.0 | 6.1 | 4.1 | 5.7 | 4.1 | 5.4 | 3.6 | 7.9 | 5.0 | 7.3 | 4.1 | 5.9 | 4.1 | 5.7 | 3.4 | 7.7 | 5.0 | 6.8 | 4.4 | 6.1 | 4.4 | 5.7 | 3.4 |
| 15 | 7.7 | 5.0 | 6.9 | 4.2 | 6.7 | 4.2 | 6.3 | 3.5 | 8.0 | 4.8 | 7.6 | 4.1 | 6.9 | 4.1 | 6.7 | 3.3 | 7.8 | 4.8 | 4.5 | 4.1 | 6.9 | 4.3 | 6.7 | 3.3 |
| 20 | 7.4 | 5.3 | 7.3 | 4.7 | 6.8 | 4.7 | 6.4 | 3.9 | 8.2 | 5.5 | 7.6 | 4.4 | 7.2 | 4.4 | 6.6 | 3.8 | 8.3 | 5.3 | 7.6 | 4.5 | 7.2 | 4.6 | 6.8 | 3.8 |
| 25 | 7.9 | 5.6 | 7.7 | 4.8 | 7.0 | 5.0 | 7.0 | 4.2 | 8.2 | 5.5 | 7.9 | 4.7 | 7.1 | 4.9 | 6.8 | 4.1 | 8.2 | 5.5 | 7.8 | 4.9 | 7.3 | 4.9 | 7.0 | 4.1 |
| 30 | 8.3 | 6.1 | 8.0 | 5.2 | 7.5 | 5.2 | 7.1 | 4.5 | 8.6 | 5.7 | 8.2 | 5.3 | 7.5 | 5.3 | 7.0 | 4.4 | 8.6 | 5.8 | 8.3 | 5.4 | 7.5 | 5.5 | 7.1 | 4.3 |

Notes : (i) 1000 iterations;
(ii) $\hat{M}_1, \hat{M}_2$, generalized spectral tests, $Q_1$ and $Q_2$ are their Chi-square approximation, respectively;
(iii) The Parzen kernel is used for both $\hat{M}_1$ and $\hat{M}_2$; $p = 10, 15, 20, 25, 30$.
(iv) A: $Y_t = 0.6Y_{t-1} + u_t$, $u_t = h_t^{1/2}\varepsilon_t, h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2$; B: $h_t = 0.2 + 0.8h_{t-1} + 0.2u_{t-1}^2$;
C: $h_t = 0.2 + 0.79h_{t-1} + 0.2u_{t-1}^2$, $\varepsilon_t \sim i.i.d.$ N(0, 1).

**TABLE 3. Empirical Sizes of Tests (cont.)**

DGP S.1: $Y_t = 0.9Y_{t-1} + u_t$, $u_t = h_t^{1/2}\varepsilon_t$

| | A: $h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2$ | | | | | | | | B: $h_t = 0.2 + 0.8h_{t-1} + 0.2u_{t-1}^2$ | | | | | | | | C: $h_t = 0.2 + 0.79h_{t-1} + 0.2u_{t-1}^2$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | |
| $p$ | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| $T = 250$ | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | 13.1 | 8.7 | 12.4 | 8.0 | 10.9 | 8.1 | 10.6 | 7.1 | 12.6 | 8.8 | 12.2 | 7.9 | 10.9 | 7.9 | 10.5 | 6.8 | 11.5 | 8.6 | 11.3 | 7.3 | 10.5 | 7.3 | 10.3 | 6.6 |
| 15 | 13.5 | 9.0 | 13.0 | 7.8 | 12.2 | 7.9 | 11.7 | 7.2 | 13.1 | 9.6 | 13.0 | 8.2 | 12.3 | 8.3 | 11.7 | 7.5 | 11.8 | 9.1 | 11.2 | 8.0 | 10.8 | 8.1 | 10.5 | 6.6 |
| 20 | 13.4 | 9.5 | 13.1 | 8.2 | 12.4 | 8.3 | 11.9 | 7.7 | 13.4 | 9.7 | 12.8 | 8.5 | 12.0 | 8.6 | 11.8 | 7.6 | 11.8 | 8.7 | 11.3 | 7.7 | 10.5 | 7.7 | 10.3 | 6.9 |
| 25 | 13.4 | 9.9 | 13.3 | 8.4 | 12.5 | 8.5 | 12.2 | 7.8 | 13.0 | 9.6 | 12.6 | 8.7 | 12.3 | 8.7 | 11.8 | 7.6 | 12.3 | 8.5 | 11.7 | 7.6 | 10.6 | 7.6 | 10.0 | 7.0 |
| 30 | 14.1 | 9.7 | 13.1 | 8.6 | 12.2 | 8.7 | 11.8 | 7.8 | 13.0 | 9.5 | 12.8 | 8.5 | 12.8 | 8.5 | 12.3 | 7.7 | 12.5 | 8.2 | 12.2 | 8.0 | 11.1 | 8.0 | 10.7 | 6.9 |
| $T = 500$ | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | 12.1 | 9.1 | 11.7 | 8.4 | 10.9 | 8.4 | 10.4 | 7.1 | 12.9 | 8.8 | 12.3 | 8.2 | 11.3 | 8.2 | 10.8 | 7.0 | 10.5 | 8.0 | 10.2 | 7.0 | 9.3 | 7.1 | 9.0 | 5.8 |
| 15 | 12.1 | 9.6 | 11.8 | 8.5 | 11.0 | 8.5 | 10.8 | 7.8 | 12.4 | 9.2 | 11.9 | 8.4 | 10.9 | 8.4 | 10.2 | 7.5 | 10.9 | 8.3 | 10.3 | 7.6 | 10.0 | 7.6 | 9.7 | 6.8 |
| 20 | 12.3 | 9.2 | 12.0 | 8.7 | 11.4 | 8.7 | 10.6 | 7.9 | 12.8 | 9.0 | 12.4 | 8.0 | 11.3 | 8.0 | 10.8 | 7.4 | 11.6 | 8.4 | 11.4 | 7.6 | 10.7 | 7.7 | 10.2 | 6.8 |
| 25 | 13.0 | 9.3 | 12.7 | 8.4 | 11.7 | 8.4 | 10.9 | 7.5 | 13.1 | 8.8 | 12.8 | 8.0 | 11.9 | 8.1 | 11.6 | 7.2 | 11.9 | 8.8 | 11.7 | 7.7 | 11.4 | 7.9 | 11.0 | 6.9 |
| 30 | 13.2 | 9.3 | 13.0 | 8.2 | 11.9 | 8.2 | 11.5 | 7.5 | 13.9 | 9.0 | 13.2 | 8.2 | 11.8 | 8.2 | 11.3 | 7.2 | 12.4 | 9.1 | 12.2 | 8.0 | 11.3 | 8.1 | 10.9 | 7.2 |
| $T = 1000$ | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | 8.5 | 5.8 | 10.6 | 7.1 | 8.9 | 6.5 | 8.9 | 6.5 | 11.0 | 7.7 | 10.4 | 6.7 | 9.4 | 6.7 | 9.0 | 5.8 | 9.5 | 6.1 | 8.9 | 5.0 | 7.8 | 5.0 | 7.3 | 3.0 |
| 15 | 9.0 | 5.2 | 10.8 | 7.6 | 9.1 | 6.4 | 9.1 | 6.4 | 10.7 | 7.4 | 10.2 | 6.5 | 9.5 | 6.5 | 9.0 | 5.5 | 9.3 | 5.6 | 9.2 | 5.2 | 7.8 | 5.3 | 7.6 | 4.1 |
| 20 | 9.4 | 5.4 | 10.7 | 7.6 | 9.9 | 6.7 | 9.9 | 6.7 | 10.8 | 6.8 | 10.5 | 5.9 | 9.9 | 6.1 | 9.4 | 5.7 | 9.3 | 5.9 | 8.4 | 5.1 | 8.1 | 5.2 | 7.8 | 4.3 |
| 25 | 9.4 | 5.7 | 10.9 | 7.9 | 9.7 | 7.0 | 9.7 | 7.0 | 10.7 | 7.6 | 10.6 | 6.3 | 10.0 | 6.3 | 9.4 | 5.6 | 8.7 | 6.2 | 8.4 | 5.3 | 7.9 | 5.5 | 7.8 | 4.8 |
| 30 | 9.7 | 6.5 | 10.5 | 7.8 | 9.8 | 7.5 | 9.8 | 7.5 | 10.6 | 7.7 | 10.5 | 6.6 | 9.9 | 6.7 | 9.4 | 6.0 | 9.1 | 6.4 | 8.6 | 5.7 | 8.1 | 5.9 | 8.1 | 5.0 |

Notes : (i) 1000 iterations;
(ii) $\hat{M}_1, \hat{M}_2$, generalized spectral tests, $Q_1$ and $Q_2$ are their Chi-square approximation, respectively;
(iii) The Parzen kernel is used for both $\hat{M}_1$ and $\hat{M}_2$; $p = 10, 15, 20, 25, 30$.
(iv) A: $Y_t = 0.9Y_{t-1} + u_t$, $u_t = h_t^{1/2}\varepsilon_t, h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2$; B: $h_t = 0.2 + 0.8h_{t-1} + 0.2u_{t-1}^2$;
C: $h_t = 0.2 + 0.79h_{t-1} + 0.2u_{t-1}^2$, $\varepsilon_t \sim i.i.d.$ N(0, 1).

## TABLE 4. Empirical Powers of Tests

| | DGP P.1: $Y_t = 0.2Y_{t-1} + 0.2Y_{t-2} + u_t$ | | | | | | | | DGP P.2: $Y_t = -0.5Y_{t-1}1(Y_{t-1} > 0) + 0.7Y_{t-1}1(Y_{t-1} \leq 0) + u_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | |
| $p$ | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| $T = 250$ | | | | | | | | | | | | | | | | |
| 10 | 61.4 | 50.8 | 61.9 | 49.8 | 61.4 | 50.8 | 62.2 | 49.9 | 98.9 | 97.2 | 98.6 | 96.6 | 98.9 | 97.1 | 98.6 | 96.8 |
| 15 | 61.3 | 51.2 | 60.0 | 49.9 | 61.3 | 51.2 | 60.3 | 49.8 | 97.8 | 94.4 | 97.2 | 92.0 | 97.8 | 94.2 | 97.4 | 92.1 |
| 20 | 60.9 | 47.6 | 58.9 | 48.3 | 60.9 | 47.6 | 59.0 | 48.6 | 96.0 | 89.4 | 94.9 | 88.5 | 96.0 | 89.3 | 95.1 | 89.1 |
| 25 | 58.9 | 47.9 | 58.0 | 45.7 | 59.0 | 47.9 | 58.0 | 46.0 | 93.9 | 86.5 | 93.2 | 84.0 | 93.8 | 86.4 | 93.2 | 84.3 |
| 30 | 58.7 | 46.6 | 57.3 | 44.1 | 58.6 | 46.6 | 57.3 | 44.3 | 92.4 | 82.3 | 91.4 | 80.1 | 92.1 | 82.0 | 91.8 | 80.6 |
| $T = 500$ | | | | | | | | | | | | | | | | |
| 10 | 89.9 | 81.4 | 89.4 | 80.0 | 89.9 | 81.4 | 89.4 | 80.1 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 15 | 89.5 | 81.9 | 89.5 | 81.6 | 89.5 | 81.9 | 89.4 | 81.6 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 89.6 | 81.4 | 89.2 | 81.1 | 89.6 | 81.4 | 89.3 | 81.1 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 25 | 89.2 | 80.2 | 89.1 | 79.7 | 89.2 | 80.2 | 89.1 | 79.7 | 100.0 | 100.0 | 100.0 | 99.8 | 100.0 | 100.0 | 100.0 | 99.8 |
| 30 | 88.4 | 79.2 | 87.8 | 79.8 | 88.4 | 79.1 | 87.9 | 79.6 | 100.0 | 99.9 | 100.0 | 99.8 | 100.0 | 100.0 | 100.0 | 99.8 |
| $T = 1000$ | | | | | | | | | | | | | | | | |
| 10 | 99.7 | 99.2 | 99.6 | 99.2 | 99.7 | 99.2 | 99.6 | 99.2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 15 | 99.6 | 99.3 | 99.6 | 99.2 | 99.6 | 99.3 | 99.6 | 99.2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 99.4 | 99.2 | 99.4 | 99.2 | 99.4 | 99.2 | 99.4 | 99.2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 25 | 99.4 | 99.1 | 99.4 | 98.9 | 99.4 | 99.1 | 99.4 | 98.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 30 | 99.4 | 99.0 | 99.4 | 98.8 | 99.4 | 99.0 | 99.4 | 98.8 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

Notes : (i) 1000 iterations;

(ii) $\hat{M}_1, \hat{M}_2$, generalized spectral tests, $Q_1$ and $Q_2$ are their Chi-square approximation, respectively;

(iii) The Parzen kernel is used for both $\hat{M}_1$ and $\hat{M}_2$; $p = 10, 15, 20, 25, 30$.

(iv) DGP P.1: $Y_t = 0.2Y_{t-1} + 0.2Y_{t-2} + u_t$, $u_t = h_t^{1/2}\varepsilon_t, h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2$;

DGP P.2: $Y_t = -0.5Y_{t-1}1(Y_{t-1} > 0) + 0.7Y_{t-1}1(Y_{t-1} \leq 0) + u_t, u_t = h_t^{1/2}\varepsilon_t, h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2$;

DGP P.3: $Y_t = 0.2Y_{t-1} + u_t, u_t = h_t^{1/2}\varepsilon_t, h_t = 0.2 + 0.6h_{t-1} + 0.1u_{t-1}^2 1(u_{t-1} > 0) + 0.5u_{t-1}^2 1(u_{t-1} \leq 0)$,
$\varepsilon_t \sim i.i.d.$ N(0, 1).

**TABLE 5. Empirical Powers of Tests (Cont.)**

| | DGP P.3: $h_t = 0.2 + 0.6h_{t-1} + 0.1u_{t-1}^2\mathbf{1}(u_{t-1} > 0) + 0.5u_{t-1}^2\mathbf{1}(u_{t-1} \leq 0)$ | | | | | | | | DGP P.4: $h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2, \{\varepsilon_t\} \sim i.i.d. \exp(1)$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | | $\hat{M}_1$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | |
| $p$ | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| | | | | | | | | $T = 250$ | | | | | | | | |
| 10 | 26.6 | 17.4 | 26.6 | 16.8 | 26.4 | 17.1 | 26.8 | 17.0 | 65.1 | 48.4 | 61.8 | 43.6 | 64.5 | 48.2 | 63.1 | 44.9 |
| 15 | 25.7 | 14.9 | 24.3 | 14.3 | 25.8 | 14.8 | 24.4 | 14.0 | 50.9 | 33.4 | 46.7 | 30.8 | 50.1 | 32.7 | 47.3 | 30.9 |
| 20 | 23.7 | 12.0 | 22.4 | 12.7 | 23.6 | 12.0 | 22.4 | 12.8 | 40.2 | 22.9 | 37.9 | 21.9 | 39.8 | 22.4 | 38.5 | 23.0 |
| 25 | 21.3 | 13.2 | 21.2 | 12.1 | 21.2 | 13.2 | 21.2 | 12.5 | 33.3 | 19.9 | 32.1 | 16.5 | 33.1 | 19.3 | 32.3 | 17.1 |
| 30 | 21.7 | 12.6 | 20.2 | 12.1 | 21.6 | 12.5 | 20.1 | 12.1 | 29.8 | 15.9 | 27.0 | 14.0 | 29.2 | 15.4 | 27.4 | 14.0 |
| | | | | | | | | $T = 500$ | | | | | | | | |
| 10 | 47.6 | 27.3 | 44.0 | 25.7 | 47.4 | 27.3 | 44.1 | 26.0 | 98.7 | 93.1 | 97.8 | 91.4 | 98.5 | 93.0 | 98.0 | 91.9 |
| 15 | 42.6 | 25.8 | 41.7 | 25.1 | 42.5 | 25.8 | 41.5 | 25.1 | 94.9 | 84.8 | 93.6 | 82.8 | 94.9 | 84.6 | 93.7 | 83.5 |
| 20 | 40.6 | 24.0 | 40.2 | 22.8 | 40.9 | 24.0 | 40.4 | 22.8 | 89.7 | 75.9 | 88.2 | 73.1 | 89.5 | 75.5 | 88.7 | 74.4 |
| 25 | 38.4 | 21.8 | 37.8 | 21.1 | 38.3 | 21.8 | 38.0 | 21.1 | 84.5 | 66.9 | 83.1 | 63.3 | 84.4 | 66.3 | 83.4 | 64.5 |
| 30 | 37.1 | 20.4 | 36.0 | 21.2 | 37.1 | 20.4 | 36.1 | 21.2 | 79.0 | 59.0 | 75.9 | 58.6 | 78.6 | 58.3 | 76.6 | 59.2 |
| | | | | | | | | $T = 1000$ | | | | | | | | |
| 10 | 80.7 | 68.5 | 78.4 | 67.0 | 80.7 | 68.2 | 78.6 | 67.3 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 15 | 78.4 | 66.2 | 76.3 | 65.3 | 78.3 | 65.8 | 76.2 | 65.5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 20 | 74.8 | 63.1 | 73.4 | 62.1 | 74.6 | 63.1 | 73.4 | 62.2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 25 | 73.7 | 61.1 | 71.6 | 57.0 | 73.7 | 61.1 | 71.9 | 57.1 | 100.0 | 100.0 | 100.0 | 99.8 | 100.0 | 100.0 | 100.0 | 100.0 |
| 30 | 72.2 | 59.1 | 70.9 | 53.5 | 72.1 | 58.8 | 71.0 | 53.5 | 100.0 | 99.6 | 100.0 | 99.5 | 100.0 | 99.6 | 100.0 | 99.6 |

Notes : (i) 1000 iterations;
(ii) $\hat{M}_1, \hat{M}_2$, generalized spectral tests, $Q_1$ and $Q_2$ are their Chi-square approximation, respectively;
(iii) The Parzen kernel is used for both $\hat{M}_1$ and $\hat{M}_2$; $p = 10, 15, 20, 25, 30$.
(iv) DGP P.3: $Y_t = 0.2Y_{t-1} + u_t, u_t = h_t^{1/2}\varepsilon_t, h_t = 0.2 + 0.6h_{t-1} + 0.1u_{t-1}^2 1(u_{t-1} > 0) + 0.5u_{t-1}^2 1(u_{t-1} \leq 0)$,
$\{\varepsilon_t\} \sim i.i.d.$ N(0, 1); DGP P.4: $Y_t = 0.2Y_{t-1} + u_t$, $u_t = h_t^{1/2}\varepsilon_t, h_t = 0.2 + 0.6h_{t-1} + 0.2u_{t-1}^2, \{\varepsilon_t\} \sim i.i.d. \exp(1)$.

**TABLE 6. Empirical Powers of Tests (Cont.)**

| | DGP P.5: | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\hat{M}_1(p)$ | | $Q_1$ | | $\hat{M}_2$ | | $Q_2$ | |
| $p$ | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| | $T = 250$ | | | | | | | |
| 10 | 88.1 | 79.7 | 86.5 | 76.0 | 87.8 | 79.2 | 87.5 | 77.0 |
| 15 | 79.6 | 69.5 | 76.3 | 65.1 | 80.9 | 72.0 | 76.7 | 66.5 |
| 20 | 71.1 | 56.5 | 68.2 | 54.9 | 70.8 | 56.3 | 68.6 | 56.3 |
| 25 | 64.0 | 51.0 | 62.7 | 46.2 | 63.1 | 50.4 | 62.6 | 47.6 |
| 30 | 58.7 | 44.4 | 55.6 | 41.0 | 57.9 | 43.4 | 56.0 | 42.3 |
| | $T = 500$ | | | | | | | |
| 10 | 99.7 | 99.5 | 99.6 | 99.3 | 99.6 | 98.2 | 99.6 | 99.3 |
| 15 | 99.5 | 98.0 | 99.3 | 98.0 | 98.9 | 96.9 | 99.3 | 97.9 |
| 20 | 98.9 | 96.6 | 98.3 | 96.5 | 97.7 | 95.8 | 98.3 | 96.5 |
| 25 | 98.0 | 95.8 | 97.7 | 94.9 | 97.1 | 94.2 | 97.7 | 94.9 |
| 30 | 96.7 | 93.7 | 96.1 | 92.9 | 96.3 | 92.0 | 96.2 | 93.1 |
| | $T = 1000$ | | | | | | | |
| 10 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 |
| 15 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 |
| 20 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 |
| 25 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 |
| 30 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 |

Notes : (i) 1000 iterations;

(ii) $\hat{M}_1, \hat{M}_2$, generalized spectral tests, $Q_1$ and $Q_2$ are their Chi-square approximation, respectively;

(iii) The Parzen kernel is used for both $\hat{M}_1$ and $\hat{M}_2$; $p = 10, 15, 20, 25, 30$.

(iv) DGP P.5: $Y_t = 0.2Y_{t-1} + u_t, u_t = h_t^{1/2}\varepsilon_t, \varepsilon_t = \frac{\exp(\lambda_t \xi_t) - \exp(0.5\lambda_t^2)}{\sqrt{\exp(2\lambda_t^2) - \exp(\lambda_t^2)}}, \{\xi_t\} \sim i.i.d.N(0,1), \lambda_t^2 = 0.2 + 0.6\lambda_{t-1}^2 + 0.2u_{t-1}^2$.