

This is a postprint version of the following published document:

Du, Z., & Escanciano, J. C. (2014). A Nonparametric Distribution-Free Test for Serial Independence of Errors. *Econometric Reviews*, 34(6-10), pp. 1011-1034.

DOI:[10.1080/07474938.2014.956616](https://doi.org/10.1080/07474938.2014.956616)

© Taylor & Francis Group



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

A Nonparametric Distribution-Free Test for Serial Independence of Errors*

ZAICHAO DU[†]

Southwestern University of Finance and Economics, China

JUAN CARLOS ESCANCIANO[‡]

Indiana University

June 7, 2013

Abstract

In this paper, we propose a test for the serial independence of unobservable errors in location-scale models. We consider a Hoeffding-Blum-Kiefer-Rosenblatt type empirical process applied to residuals, and show that under certain conditions it converges weakly to the same limit as the process based on true errors. We then consider a generalized spectral test applied to estimated residuals, and get a test that is asymptotically distribution-free and powerful against any type of pairwise dependence at all lags. Some Monte Carlo simulations validate our theoretical findings.

Keywords and Phrases: Empirical processes; Generalized spectral test; Location-scale model; Parameter estimation uncertainty; Serial dependence; Unobservable errors.

JEL Classifications: C12, C14, C52.

*We would like to thank the Editor, Esfandiari Maasoumi, and two anonymous referees for their helpful comments, which have greatly improved the paper.

[†]Corresponding Address: Research Institute of Economics and Management, Southwestern University of Finance and Economics, 55 GuangHuaCun St., Chengdu, China 610074. Email: duzc@swufe.edu.cn. Research funded by Western and Border Areas Program, Humanities and Social Sciences, Ministry of Education of China, 11XJC790002 and 211 Project for Southwestern University of Finance and Economics (3rd phase).

[‡]Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

1. INTRODUCTION

Consider the location-scale model

$$Y_t = \mu(X_t, \theta_0) + \sigma(X_t, \theta_0)u_t, \quad (1)$$

where Y_t is the dependent variable of interest, X_t is a vector of covariates, μ and σ are parametric specifications of the location and scale curves, respectively, and θ_0 is some unknown parameter in a compact set $\Theta \subset \mathbb{R}^p$.

Many models fall under this setup, to name a few, the Global Warming Trend model (Woodward and Gray, 1993, 1995), the Static Regression model, used for example in studying the relationship between permanent consumption and permanent income of individuals (see e.g. Hendry, 1995, pg. 233-241), autoregressive moving-average models with generalized autoregressive conditional heteroscedastic noises (ARMA-GARCH), widely used in financial applications, the Finite Distributed Lag model (see e.g. Hendry, 1995, pg. 273-282), and the Survival Analysis model used in Cai et al. (2008).

One important assumption for model (1) is the independent and identically distributed (*iid*) assumption on $\{u_t\}_{t \in \mathbb{Z}}$, which plays a crucial role in inferences on the parameter θ_0 and in specification tests of the parametric location and scale. In particular, the *iid* assumption often implies that the Value at Risk (*VaR*) model at level α can be expressed as

$$VaR_\alpha = \mu(X_t, \theta_0) + \sigma(X_t, \theta_0)F_u^{-1}(\alpha),$$

where $F_u^{-1}(\alpha)$ is the quantile function of u_t . *VaR* models play an important role in the assessment of market risk at commercial banks and other financial institutions.

Despite its importance, there have been only a few researches on testing the *iid* assumption of unobservable errors u_t . In this paper, we propose a new test for the null hypothesis

$$H_0 : \{u_t(\theta_0)\}_{t \in \mathbb{Z}} \text{ is a sequence of } iid \text{ r.v. for some } \theta_0 \in \Theta \subset \mathbb{R}^p, \quad (2)$$

where henceforth

$$u_t(\theta) = \frac{Y_t - \mu(X_t, \theta)}{\sigma(X_t, \theta)}, \quad \theta \in \Theta.$$

The alternative hypothesis is $H_1 : \{u_t(\theta)\}_{t \in \mathbb{Z}}$ are not independent for any $\theta \in \Theta$.

As θ_0 is unknown, we replace it with a suitable estimate $\hat{\theta}_n$ and construct residuals $\hat{u}_t = u_t(\hat{\theta}_n)$. Our test statistic is then a function of the Hoeffding-Blum-Kiefer-Rosenblatt-type empirical process (Hoeffding, 1948, Blum et al. 1961, Delgado, 1999) applied to residuals \hat{u}_t at different lags. In contrast to the general theory on empirical processes with estimated parameters, see e.g. Durbin (1973), our test is asymptotically pivotal under mild conditions. We actually show that our test

statistic based on the residuals has the same asymptotic distribution as the corresponding statistic based on the true errors.

Our arguments are based on empirical process theory under martingale difference conditions, see e.g. Delgado and Escanciano (2007). We find a necessary and sufficient condition for the estimation effect to vanish asymptotically, and we give simple sufficient conditions (see Assumption A2 and Remark 2). Our results then shed some light as to whether or not there are estimation effects in tests of serial independence of unobservable errors.

Our results generalize the findings in Delgado and Mora (2000, DM hereafter) in several directions. First, DM only considered linear regression models with fixed regressors, while here we consider a more general location-scale model (1) with stochastic regressors. Second, DM only focused on the independence between u_t and u_{t-1} , while we consider all possible lags, and our test does not involve the choice of a lag order, which is an appealing feature that most other serial dependence tests do not have.

To validate our theoretical findings, we run some Monte Carlo studies. In our simulations we find that the empirical sizes and powers of tests based on residuals are very close to those based on true errors. We then compare our results with some other tests, such as Box-Ljung-Pierce (Box and Pierce 1970, Ljung and Box 1978, BLP hereafter) test and BDS test (Brock et al. 1991, 1996). Generally, our test has better size and power performance.

The rest of the paper is organized as follows: in Section 2 we introduce our test statistic and derive its limit distribution. In Section 3 we run some Monte Carlo simulations to study the finite sample performance of the proposed test. In Section 4 we conclude and suggest some directions for future research. Mathematical proofs are gathered in the Appendix.

2. MAIN RESULTS

In the sequel, we simplify the notations as follows: $u_t = u_t(\theta_0)$ and $\hat{u}_t = u_t(\hat{\theta}_n)$, where $\hat{\theta}_n$ is a \sqrt{n} -consistent estimator for θ_0 , with n the sample size. We assume throughout that $\{u_t\}_{t \in \mathbb{Z}}$ is stationary. Let F and $F_j(x, y) = \Pr(u_t \leq x, u_{t-j} \leq y)$ denote the marginal and joint distribution functions of (u_t, u_{t-j}) , respectively; and let f be the marginal density function of u_t . Let $I(A)$ be the indicator function for the set A , and let Ω_t be the σ -algebra generated by $\{X_t, Y_{t-1}, X_{t-1}, Y_{t-2}, X_{t-2}, \dots\}$. Let C be a constant that may change from expression to expression. Finally, let Θ_0 be an arbitrary neighborhood of $\theta_0 \in \Theta$, and we assume that $\sigma(X_t, \theta)$ is bounded away from zero almost surely (a.s.) in Θ_0 .

Since we want to test the independence of the errors $\{u_t\}_{t \in \mathbb{Z}}$, we start with the following depen-

dence measure

$$\gamma_j(x, y, \theta_0) \equiv \gamma_j(x, y) = \text{Cov}[I(u_t \leq x), I(u_{t-j} \leq y)] = F_j(x, y) - F(x)F(y), \quad j \geq 0,$$

which was proposed by Hoeffding (1948). Unlike correlation, which only measures linear dependence, $\gamma_j(x, y)$ captures all types of pairwise dependence. The sample counterpart of $\gamma_j(x, y)$ based on a sample $\{u_t\}_{t=1}^n$ is

$$\gamma_{nj}(x, y) = \frac{1}{n-j} \sum_{t=1+j}^n I(u_t \leq x)I(u_{t-j} \leq y) - \frac{1}{(n-j)^2} \left\{ \sum_{t=1+j}^n I(u_t \leq x) \right\} \left\{ \sum_{t=1+j}^n I(u_{t-j} \leq y) \right\}.$$

Notice that under the null hypothesis (2),

$$\gamma_j(x, y) = 0 \quad \forall j \geq 1, \forall (x, y) \in \mathbb{R}^2.$$

Hence we can test (2) based on the distance between γ_{nj} and 0. Hoeffding (1948) used this idea to test the independence between two *iid* random variables. Along the same lines, Hong (2000) proposed a generalized spectral test for serial independence of stationary time series. However, in our present context $\{u_t\}_{t=1}^n$ is unobservable, as θ_0 is unknown. Then, we substitute \hat{u}_t for u_t in γ_{nj} , and get

$$\hat{\gamma}_{nj}(x, y) = \frac{1}{n-j} \sum_{t=1+j}^n I(\hat{u}_t \leq x)I(\hat{u}_{t-j} \leq y) - \frac{1}{(n-j)^2} \left\{ \sum_{t=1+j}^n I(\hat{u}_t \leq x) \right\} \left\{ \sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right\}.$$

In contrast to the general theory on empirical processes with estimated parameters, see e.g. Durbin (1973), we show in the next theorem that $\hat{\gamma}_{nj}(x, y)$ has the same asymptotic distribution as $\gamma_{nj}(x, y)$ under the following conditions:

Assumption A1. $\{Y_t, X_t\}_{t=1}^n$ is strictly stationary and ergodic.

Assumption A2. $E[a_t(\theta_0)|u_{t-j}] = E[a_t(\theta_0)]$ a.s. and $E[b_t(\theta_0)|u_{t-j}] = E[b_t(\theta_0)]$ a.s. $\forall j \geq 1$, where $a_t(\theta) = \dot{\mu}(X_t, \theta)/\sigma(X_t, \theta)$ and $b_t(\theta) = \dot{\sigma}(X_t, \theta)/\sigma(X_t, \theta)$, with $\dot{\mu}(X_t, \theta) = \partial\mu(X_t, \theta)/\partial\theta$ and $\dot{\sigma}(X_t, \theta) = \partial\sigma(X_t, \theta)/\partial\theta$.

Assumption A3. Under H_0 , u_t is independent of Ω_t .

Assumption A4. $f(x)$ is uniformly continuous and $\sup_{x \in \mathbb{R}} f(x) |x| < \infty$. $\mu(X_t, \theta)$ and $\sigma(X_t, \theta)$ are continuously differentiable in θ a.s. with $E[\sup_{\theta \in \Theta_0} |a_t(\theta)|] < \infty$ and $E[\sup_{\theta \in \Theta_0} |b_t(\theta)|] < \infty$.

Assumption A5. $\sqrt{n}(\widehat{\theta}_n - \theta_0) = O_p(1)$, with θ_0 in the interior of Θ .

Assumption A1 is made here for easy exposition. Our results are also valid for some non-stationary and non-ergodic sequences, see Escanciano (2007) for details. Assumption A2 is satisfied by many models, see examples outlined in the Introduction as well as the remarks after Theorem 1. Assumption A3 is often assumed in ARMA-GARCH models, among others. A3 is not needed in our arguments, but it simplifies certain quantities, see E_j in Remark 1 below. Assumption A4 is required for the asymptotic equicontinuity¹ of certain empirical processes and the uniform law of large numbers. Assumption A5 is satisfied for most estimators in the literature, such as the quasi-maximum likelihood and the generalized method of moments estimators. With these assumptions in place we are in position to establish the first important result of the paper.

Theorem 1 *Under Assumptions A1-A5 and the null hypothesis (2), we have for all $j, 1 \leq j \leq n-1$,*

$$\sup_{(x,y) \in \mathbb{R}^2} \left| \sqrt{n-j}[\widehat{\gamma}_{nj}(x,y) - \gamma_{nj}(x,y)] \right| = o_p(1). \quad (3)$$

It follows from Theorem 1 that under H_0 , $\sqrt{n-j}\widehat{\gamma}_{nj}(x,y)$ and $\sqrt{n-j}\gamma_{nj}(x,y)$ converge weakly to the same process. Therefore tests for independence of u_t based on $\widehat{\gamma}_{nj}(x,y)$ will have the same asymptotic distribution as those based on $\gamma_{nj}(x,y)$. The proof of Theorem 1 sheds some light as to whether or not there is (asymptotic) estimation effect in classical tests of independence based on $\widehat{\gamma}_{nj}(x,y)$, as shown in the following remarks.

Remark 1: *Assumption A2 is sufficient for the asymptotic equivalence under the null between $\widehat{\gamma}_{nj}(x,y)$ and $\gamma_{nj}(x,y)$, and it is also necessary for many commonly used models, including but not limited to linear and nonlinear regression models. In the proof of Theorem 1, we implicitly show that under A1, A3, A4 and A5 (i.e. without assuming A2), the following equivalence holds*

$$\sup_{(x,y) \in \mathbb{R}^2} \left| \sqrt{n-j}[\widehat{\gamma}_{nj}(x,y) - \gamma_{nj}(x,y)] - \sqrt{n}(\widehat{\theta}_n - \theta_0)' E_j(x,y) \right| = o_p(1), \quad (4)$$

where

$$E_j(x,y) = E \left[\frac{\partial \Pr[u_t(\theta_0) \leq x | \Omega_t]}{\partial \theta} \{I(u_{t-j} \leq y) - F(y)\} \right].$$

See the Appendix for details. Therefore, a necessary and sufficient condition for not having (asymptotic) estimation effect, i.e. (3), is that

$$\sup_{(x,y) \in \mathbb{R}^2} \left| \sqrt{n}(\widehat{\theta}_n - \theta_0)' E_j(x,y) \right| = o_p(1). \quad (5)$$

¹For definition of asymptotic (uniform) equicontinuity see Chapter 1.5 in van der Vaart and Wellner (1996).

One possible case for (5) to hold is that $\widehat{\theta}_n$ is superconsistent, so $\sqrt{n}(\widehat{\theta}_n - \theta_0) = o_P(1)$ (by Assumption A4 E_j is bounded). However, typically the asymptotic variance of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ exists and is positive definite. Under this mild assumption, (5) holds if and only if $E_j(x, y) \equiv 0$. Under H_0 , we have

$$\frac{\partial \Pr[u_t(\theta_0) \leq x | \Omega_t]}{\partial \theta} = f(x) (a_t(\theta_0) + x b_t(\theta_0)),$$

which jointly with A2 yield $E_j \equiv 0$. Hence, A2 is a sufficient condition for (5).

Remark 2: A sufficient condition for A2 is A2*: X_t is independent of the error u_{t-i} for any $i \geq 1$. A2* is a sufficient condition for no estimation effect but not necessary in general, as can be seen from the following example. Assume:

$$Y_t = X_t' \theta_0 + u_t, \quad X_t = u_{t-1} \cdot u_{t-2}, \quad u_t \sim iid(0, 1),$$

where obviously A2* is violated, but one can verify that $E_j \equiv 0$, and hence there is no estimation effect. Note that the fixed regressors assumption in DM implies A2*, and hence A2.

The test statistic that we propose here is an extension of Hong's (2000) generalized spectral test to estimated residuals. Define $\gamma_{-j}(\cdot) = \gamma_j(\cdot)$ for $j \geq 1$, and consider the Fourier transform of the functions $\gamma_j(\cdot)$,

$$h(\varpi, x, y) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j(x, y) e^{-ij\varpi} \quad i = \sqrt{-1}, \quad \varpi \in [-\pi, \pi], \quad (x, y) \in \mathbb{R}^2,$$

which contains the same information about the null hypothesis as $\{\gamma_j(x, y)\}_{j=0}^{\infty}$. Consider the generalized spectral distribution function

$$H(\lambda, x, y) = 2 \int_0^{\lambda\pi} h(\varpi, x, y) d\varpi = \gamma_0(x, y)\lambda + 2 \sum_{j=1}^{\infty} \gamma_j(x, y) \frac{\sin j\pi\lambda}{j\pi} \quad \lambda \in [0, 1],$$

whose sample counterpart based on residuals $\{\widehat{u}_t\}_{t=1}^n$ is

$$\widehat{H}_n(\lambda, x, y) = \widehat{\gamma}_{n0}(x, y)\lambda + 2 \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right)^{1/2} \widehat{\gamma}_{nj}(x, y) \frac{\sin j\pi\lambda}{j\pi},$$

with $(1 - j/n)^{1/2}$ a finite sample correction factor that delivers a better approximation to the finite sample distribution. Since under the null hypothesis (2), $h(\varpi, x, y) = 1/(2\pi)\gamma_0(x, y)$ and $H(\lambda, x, y) = \gamma_0(x, y)\lambda$, tests can be based on the discrepancy between $\widehat{H}_n(\lambda, x, y)$ and $\widehat{H}_{n0}(\lambda, x, y) = \widehat{\gamma}_{n0}(x, y)\lambda$, that is,

$$\widehat{S}_n(\lambda, x, y) = \left(\frac{n}{2}\right)^{1/2} [\widehat{H}_n(\lambda, x, y) - \widehat{H}_{n0}(\lambda, x, y)] = \sum_{j=1}^{n-1} (n-j)^{1/2} \widehat{\gamma}_{nj}(x, y) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi}. \quad (6)$$

Define $\Pi := [0, 1] \times \mathbb{R}^2$ and $\widehat{F}_n(x) := 1/n \sum_{t=1}^n I(\widehat{u}_t \leq x)$. To test for hypothesis H_0 , we evaluate the distance between \widehat{S}_n and 0 using the usual Cramér-von Mises norm

$$\widehat{T}_{GCM} = \int_{\Pi} (\widehat{S}_n(\lambda, x, y))^2 d\lambda d\widehat{F}_n(x) d\widehat{F}_n(y) = \sum_{j=1}^{n-1} (n-j) \widehat{\gamma}_n^2(j) \frac{1}{(j\pi)^2}, \quad (7)$$

where

$$\widehat{\gamma}_n^2(j) = n^{-2} \sum_{t=1}^n \sum_{s=1}^n \widehat{\gamma}_{nj}^2(\widehat{u}_t, \widehat{u}_s).$$

Hong (2000) proposed a generalized spectral test statistic T_{GCM} for the serial independence of raw data. Because of the distribution-free property, the exact limiting distribution of T_{GCM} can be approximated directly by simulation, see Table 1 of Hong (2000, pg. 565) for some percentiles of the distribution. Our test statistic \widehat{T}_{GCM} here extends Hong's test to estimated residuals, and we obtain the somewhat surprising result that \widehat{T}_{GCM} has the same null limit distribution as T_{GCM} in the next theorem.

Theorem 2 *Let Assumptions A1-A5 hold. Then, under the null hypothesis (2)*

$$\widehat{T}_{GCM} \xrightarrow{d} T_{\infty} = \int_{[0,1]^3} (Z(\eta))^2 d\eta,$$

where $\eta = (\lambda, x, y) \in [0, 1]^3$, and $Z(\eta)$ is a three-dimensional Brownian Bridge.²

Theorem 2 shows that \widehat{T}_{GCM} has the same null limit distribution as T_{GCM} , which can be easily simulated. Moreover, they have similar asymptotic power properties as implied by the next theorem.

Theorem 3 *Under Assumptions A1, A4 and A5, we have for all j , $1 \leq j \leq n-1$,*

$$\sup_{(x,y) \in \mathbb{R}^2} |\widehat{\gamma}_{nj}(x, y) - \gamma_{nj}(x, y)| = o_p(1).$$

Remark 3: *From Theorem 3 here and Theorem 4 of Hong (2000) it follows that \widehat{T}_{GCM} is consistent against any type of pairwise dependence. Furthermore, a similar analysis to that carried out in Hong (2000) shows that our test has non-trivial power under the class of local alternatives considered in (4.1) of Hong (2000).*

3. MONTE CARLO SIMULATIONS

To assess the finite sample performance of our proposed test, especially the effect of replacing errors by residuals, we carry out some Monte Carlo experiments. All the simulations are done

²That is, $Z(\eta)$ is a Gaussian process with mean 0 and covariance kernel $E\{Z(\eta)Z(\eta')\} = (\lambda \wedge \lambda' - \lambda\lambda')(x \wedge x' - xx')(y \wedge y' - yy')$, where $\lambda \wedge \lambda' = \min(\lambda, \lambda')$.

through the Libra High Performance Cluster at Indiana University. The software that we use is R version 2.5.1 for Windows.³

In this section we consider the following models, where $\{u_t\}$ is *iid* $N(0, 1)$:

1. *Linear*: $Y_t = X_t' \theta_0 + u_t$, where $\{X_t\}$ is an *iid* trivariate vector with first component constant, second component $N(0, 1)$ and third component $N(1, 1.5^2)$, $\theta_0 = (1, 1, 1)'$. The two normal variables are independent of each other.
2. *Exponential*: $Y_t = \exp(X_t' \theta_0) + u_t$, where X_t and θ_0 are the same as those in the linear model.
3. *Quadratic*: $Y_t = (X_t' \theta_0 - \sqrt{3})^2 + 1 + 0.2u_t$, where X_t is a trivariate vector with independent $U[0, 1]$ components, and $\theta_0 = (\sqrt{2}, \sqrt{3}, \sqrt{6})'$. This is the model considered in Xue and Zhu (2006).
4. *Partially Linear*: $Y_t = \exp(X_{1t}) + X_{2t}' \theta_0 + \sqrt{2}u_t$, where X_{1t} is $U[0, 1]$ distributed, X_{2t} is a bivariate vector with independent $N(0, 1)$ components, and $\theta_0 = (0.5, -2)'$. This is the model considered in Bravo (2007).
5. *Location Scale*: $Y_t = X_t' \theta_0 + \sqrt{\beta_0 + \beta_1 X_{2t}^2} u_t$, where X_t and θ_0 are same as those in the linear model, and $\beta_0 = \beta_1 = 1$.

For the above five models, following Hong (2000), we generate 1000 realizations of $\{Y_t, X_t\}_{t=1}^n$ with $n = 40$ and 100 . Then we estimate the parameters θ_0 in models 1-4 by the least squares method, and estimate θ_0 , β_0 and β_1 in model 5 by the maximum likelihood method. After obtaining the residuals \hat{u}_t , we calculate $\hat{\gamma}_{nj}(\hat{u}_t, \hat{u}_s)$ and \hat{T}_{GCM} given in (7).

We compare our test with the BLP and the BDS tests. Box and Pierce (1970) and Ljung and Box (1978) propose a diagnostic test for an ARMA model:

$$BLP(k) = n(n+2) \sum_{j=1}^k (n-j)^{-1} \hat{\rho}_{nj}^2(j),$$

where $\hat{\rho}_{nj}(j)$ is the lag- j sample autocorrelation function of $\{\hat{u}_t\}_{t=1}^n$. Here we only report the results for $k = 5$. We have tried other choices of k and found similar performance compared with our test. The BDS test of Brock, Hsieh and LeBaron (1991) is a popular choice for testing *iid* in regression errors. The *BDS* test statistic is computed as

$$BDS \equiv BDS(m, d) = n^{1/2} \hat{V}_m^{-1/2} \left[\hat{C}_m(d) - \left(\hat{C}_1(d) \right)^m \right],$$

based on the correlation integral (cf. Brock et al. 1991)

$$\hat{C}_m(d) = \frac{2}{n(n-1)} \sum_{t=m+1}^n \sum_{s=m}^{t-1} \prod_{j=0}^{m-1} 1(|\hat{u}_{t-j} - \hat{u}_{s-j}| < d),$$

³1000 simulations with $n = 100$ take about 5 minutes on a PC with an Intel Core 2 Duo processor and 2 GB of memory.

where m is the so-called embedding dimension, d is a distance parameter, and \widehat{V}_m is an asymptotic variance estimator. The $BDS(m, d)$ has the appealing nuisance parameter-free property that any \sqrt{n} -consistent estimator θ_n has no impact on its asymptotic null distribution under a class of conditional mean models (1) with constant variance $\sigma(X_t, \theta_0) \equiv \sigma$, cf. Brock et al. (1991). To the best of our knowledge, no extension of the later result is available for general location-scale models. We follow the recommendation of Brock et al. (1991) and use $m = 2$ and $d = \widehat{\sigma}$, where $\widehat{\sigma}$ is the sample standard deviation of residuals. We have examined other choices of m and d leading to similar conclusions.

Table 1 gives the empirical sizes of \widehat{T}_{GCM} for the above models when sample size $n = 40$. Since we have the true errors $\{u_t\}_{t=1}^{40}$ in the simulations, we can also calculate the test statistic T_{GCM} (Hong 2000), which is defined in the same way as \widehat{T}_{GCM} except using the true errors instead of the residuals, i.e.

$$T_{GCM} = \sum_{j=1}^{n-1} (n-j) \gamma_n^2(j) \frac{1}{(j\pi)^2},$$

where

$$\gamma_n^2(j) = n^{-2} \sum_{t=1}^n \sum_{s=1}^n \gamma_{nj}^2(u_t, u_s).$$

The critical values that we use for \widehat{T}_{GCM} and T_{GCM} are 0.006598 for the nominal level $\alpha = 0.1$ and 0.007675 for $\alpha = 0.05$, which are taken from Hong (2000, pg. 565). We can see that \widehat{T}_{GCM} and T_{GCM} have similar empirical sizes that are fairly close to the nominal levels, which supports our theoretical findings in Theorem 2: \widehat{T}_{GCM} has the same null limit distribution as T_{GCM} . As comparison, we also include the sizes of BLP and BDS tests. BLP has sizes comparable with our test, while BDS has much worse sizes than our test. For $n = 100$, we get similar results, and we do not report them here to save space.

Table 1. Size of \widehat{T}_{GCM} , T_{GCM} , BLP and BDS at 10%, 5%, $n = 40$

Models	$\alpha = 10\%$				$\alpha = 5\%$			
	\widehat{T}_{GCM}	T_{GCM}	BLP	BDS	\widehat{T}_{GCM}	T_{GCM}	BLP	BDS
<i>Linear</i>	9.1%	9.4%	11.1%	34.2%	4.5%	4.9%	5.7%	26.7%
<i>Exponential</i>	8.8%	9.4%	10.6%	33.5%	4.4%	4.9%	5.9%	27.7%
<i>Quadratic</i>	10.3%	10.8%	10.5%	36.1%	5.8%	5.6%	6.3%	29.5%
<i>Partially Linear</i>	9.2%	9.5%	9.0%	36.8%	4.6%	5.0%	5.7%	29.4%
<i>Location Scale</i>	9.1%	9.4%	11.6%	39.6%	4.8%	4.9%	6.2%	30.1%

To investigate the power, we first consider alternatives where $\{u_t\}$ are not *iid* due to the incorrect specification of the conditional mean and variance. Hereinafter $\{\varepsilon_t\}$ is *iid* $N(0, 1)$.

ALT1. *AR-Mean*: $Y_t = 0.2Y_{t-1} + \mu(X_t, \theta_0) + \sigma(X_t, \theta_0)\varepsilon_t$, where $\mu(X_t, \theta_0)$ and $\sigma^2(X_t, \theta_0)$ are the conditional mean and variance in the null models, respectively.⁴ Here we have

$$u_t = \frac{0.2[Y_{t-1} - E(Y_t)] + \sigma(X_t, \theta_0)\varepsilon_t}{\sqrt{0.2^2 \text{var}(Y_t) + \sigma^2(X_t, \theta_0)}}.$$

ALT2. *AR-GARCH*: $Y_t = 0.2Y_{t-1} + \mu(X_t, \theta_0) + \sigma_t\varepsilon_t$, where $\mu(X_t, \theta_0)$ is as in ALT1, and $\sigma_t^2 = 0.05 + 0.1\sigma_{t-1}^2\varepsilon_{t-1}^2 + 0.85\sigma_{t-1}^2$. Here

$$u_t = \frac{0.2[Y_{t-1} - E(Y_t)] + \sigma_t\varepsilon_t}{\sqrt{0.2^2 \text{var}(Y_t) + E(\sigma_t^2)}}.$$

We then follow the designs in Hong (2000) and consider the following alternatives for the errors u_t :

ALT3. *AR(1)*: $u_t = 0.3u_{t-1} + \varepsilon_t$.

ALT4. *Bilinear*: $u_t = \varepsilon_t(0.2 + 0.5u_{t-1})$.

ALT5. *Non-linear moving average*: $u_t = \varepsilon_{t-1}(0.8 + \varepsilon_t)$.

ALT6. *TAR(1)*: $u_t = -0.5u_{t-1} + \varepsilon_t$, if $u_{t-1} \leq 1$, and $u_t = 0.4u_{t-1} + \varepsilon_t$ otherwise.

For each alternative, we generate 1000 realizations of $\{Y_t, X_t\}_{t=1}^{100}$. The models fitted are the same models under the null, 1 to 5 above. The size-corrected power of \hat{T}_{GCM} at the 5% level is reported in Table 2.⁵ As we have the true errors $\{u_t\}_{t=1}^{100}$, we can also calculate T_{GCM} , whose power at 5% level is included for the sake of comparison. We also include the powers of BLP and BDS tests.

We observe that \hat{T}_{GCM} and T_{GCM} have both high empirical powers. Their power is comparable for all alternatives. Also notice that even for alternatives with zero autocorrelations (*Bilinear*, *Non-linear moving average* and *TAR(1)*) our test still has good power, while BLP has no power against such alternatives. BDS has worse power than \hat{T}_{GCM} in most cases except two occasions.

The above simulations support our findings on lack of estimation effects. To test H_0 for those cases, we can use the estimated residuals to construct the test statistic \hat{T}_{GCM} , which will have the same asymptotic distribution as the test based on the true errors T_{GCM} .

Next, to show that there are estimation effects in general when Assumption A2 is violated, we run some simulations for the simple AR(1) Model: $Y_t = \theta_0 Y_{t-1} + u_t$, where $|\theta_0| < 1$ and $u_t \sim iid(0, 1)$. For this model we can show that A2 is a necessary and sufficient condition for no asymptotic estimation

⁴For example $\mu(X_t, \theta_0) = (X_t' \theta_0 - \sqrt{3})^2 + 1$ and $\sigma(X_t, \theta_0) = 0.2$ for the Quadratic model.

⁵As implied in Hong (2000), $\{u_t\}$ in these alternatives are α -strong mixing. One can also verify that the long-run variance of $X_t u_t$ is finite. Therefore the Central Limit Theorem for stationary and α -mixing process can be applied to $X_t u_t$. As our estimation methods here are either least square method or Gaussian MLE, Assumption A5 also holds for these alternatives.

effect. It is easy to verify here that $a_t(\theta_0) = Y_{t-1}$, and hence A2 is violated. To illustrate the estimation effect, we consider $\theta_0 = -0.8, -0.4, 0.4, 0.8$. Table 3 and 4 display the sizes and power of T_{GCM} and \hat{T}_{GCM} , respectively.

One can see that due to the presence of the estimation effect in this case \hat{T}_{GCM} no longer has similar sizes as T_{GCM} . The difference is quite obvious when compared with Table 1 and 2.

Table 2. Size-corrected power of \hat{T}_{GCM} , T_{GCM} , BLP and BDS at 5%, $n = 100$

Models	<i>AR-Mean</i>				<i>AR-GARCH</i>			
	\hat{T}_{GCM}	T_{GCM}	<i>BLP</i>	<i>BDS</i>	\hat{T}_{GCM}	T_{GCM}	<i>BLP</i>	<i>BDS</i>
<i>Linear</i>	.356	.354	.212	.128	.350	.347	.259	.241
<i>Exponential</i>	.935	.974	.108	.197	.937	.964	.084	.182
<i>Quadratic</i>	.380	.377	.218	.227	.382	.361	.237	.322
<i>Partially Linear</i>	.339	.377	.213	.121	.331	.370	.231	.233
<i>Location Scale</i>	.400	.564	.245	.150	.363	.360	.276	.236
Models	<i>AR(1)</i>				<i>Bilinear</i>			
	\hat{T}_{GCM}	T_{GCM}	<i>BLP</i>	<i>BDS</i>	\hat{T}_{GCM}	T_{GCM}	<i>BLP</i>	<i>BDS</i>
<i>Linear</i>	.694	.705	.535	.280	.746	.785	.073	.271
<i>Exponential</i>	.686	.705	.541	.277	.735	.785	.079	.259
<i>Quadratic</i>	.701	.711	.519	.283	.719	.792	.096	.298
<i>Partially Linear</i>	.695	.712	.537	.310	.697	.780	.091	.285
<i>Location Scale</i>	.686	.705	.487	.287	.710	.785	.065	.174
Models	<i>Nonlinear</i>				<i>TAR(1)</i>			
	\hat{T}_{GCM}	T_{GCM}	<i>BLP</i>	<i>BDS</i>	\hat{T}_{GCM}	T_{GCM}	<i>BLP</i>	<i>BDS</i>
<i>Linear</i>	.424	.449	.062	.361	.426	.440	.050	.307
<i>Exponential</i>	.400	.449	.074	.362	.375	.440	.052	.295
<i>Quadratic</i>	.372	.434	.068	.403	.384	.423	.046	.360
<i>Partially Linear</i>	.344	.423	.080	.408	.364	.446	.072	.344
<i>Location Scale</i>	.387	.449	.065	.357	.400	.440	.054	.311

Table 3. Size of \widehat{T}_{GCM} and T_{GCM} for AR(1), $n = 40$

θ_0	$\alpha = 10\%$		$\alpha = 5\%$	
	\widehat{T}_{GCM}	T_{GCM}	\widehat{T}_{GCM}	T_{GCM}
-0.8	5.7%	9.9%	2.4%	4.4%
-0.4	2.1%	9.9%	0.6%	4.4%
0.4	2.1%	9.9%	0.8%	4.4%
0.8	5.9%	9.9%	1.8%	4.4%

Table 4. Power of \widehat{T}_{GCM} and T_{GCM} at 5% for AR(1), $n = 40$

θ_0	<i>AR(1)</i>		<i>Bilinear</i>		<i>Nonlinear</i>		<i>TAR(1)</i>	
	\widehat{T}_{GCM}	T_{GCM}	\widehat{T}_{GCM}	T_{GCM}	\widehat{T}_{GCM}	T_{GCM}	\widehat{T}_{GCM}	T_{GCM}
-0.8	.064	.339	.174	.284	.120	.190	.097	.199
-0.4	.007	.339	.145	.284	.079	.190	.050	.199
0.4	.013	.339	.125	.284	.079	.190	.052	.199
0.8	.155	.339	.169	.284	.117	.190	.089	.199

4. CONCLUSIONS

In this paper, we propose a test for the serial independence of errors u_t in the location-scale model $Y_t = \mu(X_t, \theta_0) + \sigma(X_t, \theta_0)u_t$. We show that, provided some conditions are satisfied, test statistics based on the estimated residuals have the same asymptotic distributions as the corresponding statistics based on the true errors, which extends the findings in DM to a much broader and more realistic context. We also extend Hong's (2000) generalized spectral test to estimated residuals, and provide a test that is asymptotically distribution-free and powerful against pairwise dependence at all lags. Some Monte Carlo simulations support our theoretical findings.

We point out some possible directions for future work. First, when estimation effects are present, the limiting distribution of our test is not pivotal, and bootstrap or related methods need to be used. These methods will be developed in future research. Although we focus on univariate u_t in this paper, we can easily extend our results to the multivariate case. When the dimension of u_t is large, we may want to use sample dependence measures based on the empirical characteristic function instead of the empirical cumulative distribution function because of the curse of dimensionality. Another possible extension is to apply the methodology here to testing the independence between u_t and X_t .

5. APPENDIX

We first develop some preliminary results that will be used in the proofs of our main theorems.

We define the process

$$K_n(c, x, y) = \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \left\{ I \left(u_t \left(\theta_0 + \frac{c}{\sqrt{n}} \right) \leq x \right) - E \left[I \left(u_t \left(\theta_0 + \frac{c}{\sqrt{n}} \right) \leq x \right) \middle| \Omega_t \right] \right\} I \left(u_{t-j} \left(\theta_0 + \frac{c}{\sqrt{n}} \right) \leq y \right),$$

indexed by $(c, x, y) \in C_K \times [-\infty, \infty]^2$, where $C_K = \{c \in \mathbb{R}^p : |c| \leq K\}$, and $K > 0$ is an arbitrary but fixed constant. Let d_K be the metric on C_K induced by the Euclidean norm, and define a metric d_R on $[-\infty, \infty]$ by $d_R(x_1, x_2) = |F(x_1) - F(x_2)|$. Then a metric d on $C_K \times [-\infty, \infty]^2$ can be defined as $d((c_1, x_1, y_1), (c_2, x_2, y_2)) = \max(d_K(c_1, c_2), d_R(x_1, x_2), d_R(y_1, y_2))$. We can see that $C_K \times [-\infty, \infty]^2$ is compact with respect to d . For definition of asymptotic uniform d -equicontinuity see Chapter 1.5 in van der Vaart and Wellner (1996).

LEMMA A1: *Under Assumptions A1 and A4, $K_n(\cdot)$ is asymptotically uniform d -equicontinuous.*

PROOF OF LEMMA A1: The result follows from the same steps of Lemma A1 in Escanciano and Olmo (2010). ■

LEMMA A2: *Under Assumptions A1 and A4, we have $\sup_{(x,y) \in \mathbb{R}^2} |K_n(\hat{c}, x, y) - K_n(0, x, y)| = o_p(1)$, for any $\hat{c} = O_p(1)$.*

PROOF OF LEMMA A2: Let $\varepsilon, \varsigma > 0$ be given. We want to show that

$$P \left(\sup_{(x,y) \in \mathbb{R}^2} |K_n(\hat{c}, x, y) - K_n(0, x, y)| > \varsigma \right) < \varepsilon \quad \text{as } n \rightarrow \infty.$$

Since $\hat{c} = O_p(1)$, there exists $K > 0$ such that $P(|\hat{c}| > K) < \varepsilon/2$.

Next, we show that for any fixed $(c, x, y) \in C_K \times \mathbb{R}^2$

$$E[|K_n(c, x, y) - K_n(0, x, y)|^2] = o(1). \tag{8}$$

To simplify notations, define

$$\begin{aligned} I_t(c, x) & : = I \left(u_t \left(\theta_0 + \frac{c}{\sqrt{n}} \right) \leq x \right), \\ IDE_t(c, x) & : = I_t(c, x) - E[I_t(c, x) | \Omega_t]. \end{aligned}$$

Assumption A1 and the martingale difference sequence property of $K_n(c, x, y)$ imply that

$$\begin{aligned}
E[|K_n(c, x, y) - K_n(0, x, y)|^2] &= E \left[|IDE_t(c, x)I_{t-j}(c, y) - IDE_t(0, x)I_{t-j}(0, y)|^2 \right] \\
&\leq 2E [|IDE_t(c, x)I_{t-j}(c, y) - IDE_t(0, x)I_{t-j}(0, y)|] \\
&\leq 2E [|IDE_t(c, x)I_{t-j}(c, y) - IDE_t(c, x)I_{t-j}(0, y)|] \\
&\quad + 2E [|IDE_t(c, x)I_{t-j}(0, y) - IDE_t(0, x)I_{t-j}(0, y)|] \\
&\leq 2E [|E[I_{t-j}(c, y)|\Omega_{t-j}] - E[I_{t-j}(0, y)|\Omega_{t-j}]|] \\
&\quad + 4E [|E[I_t(c, y)|\Omega_t] - E[I_t(0, y)|\Omega_t]|] \\
&= O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

The last equality follows from the Mean Value Theorem and Assumption A4, see the proof of Theorem 1 for further details.

Equation (8) and Chebyshev's Inequality imply that

$$|K_n(c, x, y) - K_n(0, x, y)| = o_p(1) \quad \forall (c, x, y) \in C_K \times \mathbb{R}^2.$$

The above display (i.e. convergence of the finite-dimensional distributions), the compactness of $C_K \times [-\infty, \infty]^2$ with respect to d and the asymptotic uniform d -equicontinuity of $K_n(c, x, y)$ imply

$$P\left(\sup_{(c, x, y) \in C_K \times \mathbb{R}^2} |K_n(c, x, y) - K_n(0, x, y)| > \varsigma\right) \leq \frac{\varepsilon}{2},$$

for n sufficiently large (see e.g. Theorem 2.1 in Newey 1991). Therefore, for n sufficiently large

$$\begin{aligned}
P\left(\sup_{(x, y) \in \mathbb{R}^2} |K_n(\hat{c}, x, y) - K_n(0, x, y)| > \varsigma\right) &\leq P(|\hat{c}| > K) + \\
&\quad P\left(\sup_{(c, x, y) \in C_K \times \mathbb{R}^2} |K_n(c, x, y) - K_n(0, x, y)| > \varsigma\right) \\
&< \varepsilon.
\end{aligned}$$

■

LEMMA A3: *Under Assumptions A1, A3 and A4, we have*

$$\sup_{(x, y) \in \mathbb{R}^2} \left| \frac{1}{n} \sum_{t=1+j}^n \frac{\partial F_t(\tilde{\theta}_n, x)}{\partial \theta} I(\hat{u}_{t-j} \leq y) - E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} I(u_{t-j} \leq y) \right] \right| = o_p(1),$$

for any consistent estimate $\tilde{\theta}_n$ of θ_0 , where henceforth

$$F_t(\theta, x) = \Pr[u_t(\theta) \leq x | \Omega_t]. \quad (9)$$

PROOF OF LEMMA A3: Notice that

$$\begin{aligned}
F_t(\theta, x) &= \Pr[u_t(\theta) \leq x | \Omega_t] \\
&= \Pr \left[\frac{Y_t - \mu(X_t, \theta)}{\sigma(X_t, \theta)} \leq x \mid \Omega_t \right] \\
&= \Pr \left[u_t \leq \frac{\mu(X_t, \theta) - \mu(X_t, \theta_0) + x\sigma(X_t, \theta)}{\sigma(X_t, \theta_0)} \mid \Omega_t \right].
\end{aligned}$$

Under Assumption A3 we have

$$\begin{aligned}
F_t(\theta, x) &= F \left(\frac{\mu(X_t, \theta) - \mu(X_t, \theta_0) + x\sigma(X_t, \theta)}{\sigma(X_t, \theta_0)} \right), \\
F_t(\theta_0, x) &= F(x), \\
\frac{\partial F_t(\theta_0, x)}{\partial \theta} &= f(x) \left(\frac{\dot{\mu}(X_t, \theta_0)}{\sigma(X_t, \theta_0)} + x \frac{\dot{\sigma}(X_t, \theta_0)}{\sigma(X_t, \theta_0)} \right) \\
&= f(x)a_t(\theta_0) + f(x)xb_t(\theta_0).
\end{aligned} \tag{10}$$

Define a metric d on $\Theta_0^2 \times [-\infty, \infty]^2$ by $d((\theta_{11}, \theta_{21}, x_1, y_1), (\theta_{12}, \theta_{22}, x_2, y_2)) = |\theta_{11} - \theta_{12}| + |\theta_{21} - \theta_{22}| + |F(x_1) - F(x_2)| + |F(y_1) - F(y_2)|$, and then $\Theta_0^2 \times [-\infty, \infty]^2$ is compact with respect to d . By Assumption A4, for each $(\theta_1, \theta_2, x, y) \in \Theta_0^2 \times \mathbb{R}^2$, $\partial F_t(\theta_1, x) / \partial \theta I(u_{t-j}(\theta_2) \leq y)$ is discontinuous at $\{u_{t-j}(\theta_2) = y\}$, which occurs with probability 0, and moreover

$$E \left[\sup_{(\theta_1, \theta_2, x, y) \in \Theta_0^2 \times \mathbb{R}^2} \left| \frac{\partial F_t(\theta_1, x)}{\partial \theta} I(u_{t-j}(\theta_2) \leq y) \right| \right] < \infty.$$

Therefore, we can apply the uniform law of larger number (ULLN) of Newey and McFadden (1994), Lemma 2.4, to the empirical process

$$\frac{1}{n} \sum_{t=1+j}^n \frac{\partial F_t(\theta_1, x)}{\partial \theta} I(u_{t-j}(\theta_2) \leq y), \quad \theta_1 \in \Theta_0, \theta_2 \in \Theta_0, x \in \mathbb{R}, y \in \mathbb{R},$$

and conclude

$$\sup_{(\theta_1, \theta_2, x, y) \in \Theta_0^2 \times \mathbb{R}^2} \left| \frac{1}{n} \sum_{t=1+j}^n \frac{\partial F_t(\theta_1, x)}{\partial \theta} I(u_{t-j}(\theta_2) \leq y) - E \left[\frac{\partial F_t(\theta_1, x)}{\partial \theta} I(u_{t-j}(\theta_2) \leq y) \right] \right| = o_p(1).$$

Then, since $\tilde{\theta}_n$ and $\hat{\theta}_n$ are consistent estimates of θ_0 and the mapping $\theta \rightarrow E[\partial F_t(\theta, x) / \partial \theta I(u_{t-j}(\theta) \leq y)]$ is continuous, we obtain

$$\begin{aligned}
&\sup_{(x, y) \in \mathbb{R}^2} \left| \frac{1}{n} \sum_{t=1+j}^n \frac{\partial F_t(\tilde{\theta}_n, x)}{\partial \theta} I(\hat{u}_{t-j} \leq y) - E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} I(u_{t-j} \leq y) \right] \right| \\
&\leq \sup_{(x, y) \in \mathbb{R}^2} \left| \frac{1}{n} \sum_{t=1+j}^n \frac{\partial F_t(\tilde{\theta}_n, x)}{\partial \theta} I(\hat{u}_{t-j} \leq y) - E \left[\frac{\partial F_t(\tilde{\theta}_n, x)}{\partial \theta} I(\hat{u}_{t-j} \leq y) \right] \right| \\
&\quad + \sup_{(x, y) \in \mathbb{R}^2} \left| E \left[\frac{\partial F_t(\tilde{\theta}_n, x)}{\partial \theta} I(\hat{u}_{t-j} \leq y) \right] - E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} I(u_{t-j} \leq y) \right] \right| \\
&= o_p(1).
\end{aligned}$$

■

In the sequel, we simplify notations as follows: $\widehat{F}_{j,n}(x) = 1/(n-j) \sum_{t=1+j}^n I(\widehat{u}_t \leq x)$, $\widehat{F}_{1,n-j}(y) = 1/(n-j) \sum_{t=1+j}^n I(\widehat{u}_{t-j} \leq y)$, $F_{j,n}(x) = 1/(n-j) \sum_{t=1+j}^n I(u_t \leq x)$ and $F_{1,n-j}(y) = 1/(n-j) \sum_{t=1+j}^n I(u_{t-j} \leq y)$.

PROOF OF THEOREM 1: We can write

$$\begin{aligned} \sqrt{n-j} \{\widehat{\gamma}_{nj}(x, y) - \gamma_{nj}(x, y)\} &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{I(\widehat{u}_t \leq x)I(\widehat{u}_{t-j} \leq y) - I(u_t \leq x)I(u_{t-j} \leq y)\} - \\ &\quad \sqrt{n-j} \left[\widehat{F}_{j,n}(x)\widehat{F}_{1,n-j}(y) - F_{j,n}(x)F_{1,n-j}(y) \right] \\ &\equiv A_{1nj}(x, y) - A_{2nj}(x, y). \end{aligned}$$

To handle A_{1nj} , we apply Lemma A2 to $\widehat{c} = \sqrt{n}(\widehat{\theta}_n - \theta_0) = O_p(1)$ and get

$$\sup_{(x,y) \in \mathbb{R}^2} \frac{1}{\sqrt{n-j}} \left| \sum_{t=1+j}^n \{I(\widehat{u}_t \leq x) - E[I(\widehat{u}_t \leq x)|\Omega_t]\}I(\widehat{u}_{t-j} \leq y) - \sum_{t=1+j}^n \{I(u_t \leq x) - E[I(u_t \leq x)|\Omega_t]\}I(u_{t-j} \leq y) \right| = o_p(1).$$

Hence,

$$\begin{aligned} A_{1nj}(x, y) &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{I(\widehat{u}_t \leq x)I(\widehat{u}_{t-j} \leq y) - I(u_t \leq x)I(u_{t-j} \leq y)\} \\ &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{E[I(\widehat{u}_t \leq x)|\Omega_t] - E[I(u_t \leq x)|\Omega_t]\}I(\widehat{u}_{t-j} \leq y) + \\ &\quad \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n E[I(u_t \leq x)|\Omega_t]\{I(\widehat{u}_{t-j} \leq y) - I(u_{t-j} \leq y)\} + o_p(1) \\ &\equiv B_{1nj}(x, y) + B_{2nj}(x, y) + o_p(1). \end{aligned}$$

Using the notation in (9), we have $E[I(\widehat{u}_t \leq x)|\Omega_t] = F_t(\widehat{\theta}_n, x)$, $E[I(u_t \leq x)|\Omega_t] = F_t(\theta_0, x)$.

Applying these notations and the Mean Value Theorem, we get

$$\begin{aligned} B_{1nj}(x, y) &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{F_t(\widehat{\theta}_n, x) - F_t(\theta_0, x)\}I(\widehat{u}_{t-j} \leq y) \\ &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n (\widehat{\theta}_n - \theta_0)' \frac{\partial F_t(\widetilde{\theta}_n, x)}{\partial \theta} I(\widehat{u}_{t-j} \leq y) \\ &= \sqrt{n}(\widehat{\theta}_n - \theta_0)' \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \frac{\partial F_t(\widetilde{\theta}_n, x)}{\partial \theta} I(\widehat{u}_{t-j} \leq y) \\ &= \sqrt{n}(\widehat{\theta}_n - \theta_0)' E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} I(u_{t-j} \leq y) \right] + o_p(1) \\ &= \sqrt{n}(\widehat{\theta}_n - \theta_0)' F(y) E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} \right] + o_p(1), \end{aligned}$$

where the second to last equality follows from Lemma A3. The last equality follows from (10) and Assumption A2.

Similarly,

$$\begin{aligned} B_{2nj}(x, y) &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n F_t(\theta_0, x) \{I(\hat{u}_{t-j} \leq y) - I(u_{t-j} \leq y)\} \\ &= F(x) \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{I(\hat{u}_{t-j} \leq y) - I(u_{t-j} \leq y)\}. \end{aligned}$$

Define the process

$$K_n(c, y) = \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \left\{ I\left(u_{t-j}\left(\theta_0 + \frac{c}{\sqrt{n}}\right) \leq y\right) - E\left[I\left(u_{t-j}\left(\theta_0 + \frac{c}{\sqrt{n}}\right) \leq y\right) \middle| \Omega_{t-j}\right] \right\},$$

indexed by $(c, y) \in C_K \times \mathbb{R}$, where $C_K = \{c \in \mathbb{R}^p : |c| \leq K\}$, and $K > 0$ is an arbitrary but fixed constant. Using the same argument as we use for A_{1nj} , we get

$$\begin{aligned} B_{2nj}(x, y) &= F(x) \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{F_{t-j}(\hat{\theta}_n, y) - F_{t-j}(\theta_0, y)\} + o_p(1) \\ &= F(x) \sqrt{n}(\hat{\theta}_n - \theta_0)' \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \frac{\partial F_{t-j}(\tilde{\theta}_n, y)}{\partial \theta} + o_p(1) \\ &= \sqrt{n}(\hat{\theta}_n - \theta_0)' F(x) E\left[\frac{\partial F_t(\theta_0, y)}{\partial \theta}\right] + o_p(1). \end{aligned}$$

Now, we turn to A_{2nj} and write

$$\begin{aligned} A_{2nj}(x, y) &= \sqrt{n-j} \left[\hat{F}_{j,n}(x) \hat{F}_{1,n-j}(y) - F_{j,n}(x) F_{1,n-j}(y) \right] \\ &= \sqrt{n-j} \left[\hat{F}_{j,n}(x) \hat{F}_{1,n-j}(y) - F_{j,n}(x) \hat{F}_{1,n-j}(y) \right] + \\ &\quad \sqrt{n-j} \left[F_{j,n}(x) \hat{F}_{1,n-j}(y) - F_{j,n}(x) F_{1,n-j}(y) \right] \\ &\equiv C_{1nj}(x, y) + C_{2nj}(x, y). \\ C_{1nj}(x, y) &= \sqrt{n-j} \left[\hat{F}_{j,n}(x) \hat{F}_{1,n-j}(y) - F_{j,n}(x) \hat{F}_{1,n-j}(y) \right] \\ &= \hat{F}_{1,n-j}(y) \frac{1}{\sqrt{n-j}} \left\{ \sum_{t=1+j}^n [I(\hat{u}_t \leq x) - I(u_t \leq x)] \right\}. \end{aligned}$$

Using the same argument as we use for B_{2n} , we get

$$\begin{aligned} C_{1nj}(x, y) &= [F(y) + o_p(1)] \left\{ \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n [F_t(\hat{\theta}_n, x) - F_t(\theta_0, x)] + o_p(1) \right\} \\ &= \sqrt{n}(\hat{\theta}_n - \theta_0)' F(y) \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \frac{\partial F_t(\tilde{\theta}_n, x)}{\partial \theta} + o_p(1) \\ &= \sqrt{n}(\hat{\theta}_n - \theta_0)' F(y) E\left[\frac{\partial F_t(\theta_0, x)}{\partial \theta}\right] + o_p(1). \end{aligned}$$

Similarly,

$$\begin{aligned}
C_{2nj}(x, y) &= \sqrt{n-j} \left[F_{j,n}(x) \widehat{F}_{1,n-j}(y) - F_{j,n}(x) F_{1,n-j}(y) \right] \\
&= F_{j,n}(x) \frac{1}{\sqrt{n-j}} \left\{ \sum_{t=1+j}^n [I(\widehat{u}_{t-j} \leq y) - I(u_{t-j} \leq y)] \right\} \\
&= \sqrt{n}(\widehat{\theta}_n - \theta_0)' F(x) E \left[\frac{\partial F_t(\theta_0, y)}{\partial \theta} \right] + o_p(1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sup_{(x,y) \in \mathbb{R}^2} \left| \sqrt{n-j} [\widehat{\gamma}_{nj}(x, y) - \gamma_{nj}(x, y)] \right| &= \sup_{(x,y) \in \mathbb{R}^2} |A_{1nj}(x, y) - A_{2nj}(x, y)| \\
&= \sup_{(x,y) \in \mathbb{R}^2} |B_{1nj}(x, y) + B_{2nj}(x, y) - \\
&\quad C_{1nj}(x, y) - C_{2nj}(x, y) + o_p(1)| \\
&= o_p(1).
\end{aligned}$$

This completes the proof of Theorem 1. ■

Next, we derive equation (4) without imposing Assumption A2.

PROOF OF EQUATION (4): By (10), we have

$$E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} I(u_{t-j} \leq y) \right] = f(x) E [a_t(\theta_0) I(u_{t-j} \leq y)] + f(x) x E [b_t(\theta_0) I(u_{t-j} \leq y)]$$

If without Assumption A2, then in general we have

$$E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} I(u_{t-j} \leq y) \right] \neq E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} \right] F(y).$$

Hence

$$B_{1nj}(x, y) = \sqrt{n}(\widehat{\theta}_n - \theta_0)' E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} I(u_{t-j} \leq y) \right] + o_p(1),$$

while the expressions for $B_{2nj}(x, y)$, $C_{1nj}(x, y)$ and $C_{2nj}(x, y)$ remain the same as above. Collecting the four terms, we get expression (4).

■

PROOF OF THEOREM 2: Recall that $\Pi = [0, 1] \times \mathbb{R}^2$ and $\eta = (\lambda, x, y) \in \Pi$. Let ν be the product measure of the Lebesgue measure on $[0, 1]$ and $F(\cdot) \times F(\cdot)$, i.e. $d\nu(\eta) = d\lambda dF(x) dF(y)$. Let $L_2(\Pi, \nu)$ be the Hilbert space of all real-valued and square ν -integrable functions on Π , furnished with the inner-product $\langle f, g \rangle = \int_{\Pi} f(\eta) g(\eta) d\nu(\eta)$. The space $L_2(\Pi, \nu)$ is endowed with the natural Borel σ -field induced by the norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$. See e.g. Bosq (2000) for a review of random elements taking values in a Hilbert space. With these notations, we have $T_{GCM} = \|S_n(\eta)\|^2 + o_p(1)$, where

$$S_n(\eta) := S_n(\lambda, x, y) = \sum_{j=1}^{n-1} (n-j)^{1/2} \gamma_{nj}(x, y) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi}.$$

Using the same arguments as in Theorem 1, we shall prove that in $L_2(\Pi, v)$,

$$\widehat{S}_n(\eta) = S_n(\eta) + o_p(1),$$

where $\widehat{S}_n(\eta)$ is defined in (6). Using the notations A_{1nj} and A_{2nj} defined in Theorem 1, we have

$$\begin{aligned}\widehat{S}_n(\eta) - S_n(\eta) &= \sum_{j=1}^{n-1} A_{1nj}(x, y) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi} - \sum_{j=1}^{n-1} A_{2nj}(x, y) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi} \\ &\equiv A_{1n}(x, y, \lambda) - A_{2n}(x, y, \lambda).\end{aligned}$$

We handle A_{1n} and A_{2n} in the same way as we handle A_{1nj} and A_{2nj} in Theorem 1.

Simple algebra shows that

$$\begin{aligned}& \sum_{j=1}^{n-1} (n-j)^{-1/2} \left(\sum_{t=1+j}^n \{I(\widehat{u}_t \leq x) - E[I(\widehat{u}_t \leq x)|\Omega_t]\} I(\widehat{u}_{t-j} \leq y) \right) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi} \\ &= \frac{1}{\sqrt{n}} \sum_{t=2}^n \{I(\widehat{u}_t \leq x) - E[I(\widehat{u}_t \leq x)|\Omega_t]\} \left(\sum_{j=1}^{t-1} \left(\frac{n}{n-j} \right)^{1/2} I(\widehat{u}_{t-j} \leq y) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi} \right) \\ &\equiv \frac{1}{\sqrt{n}} \sum_{t=2}^n \{I(\widehat{u}_t \leq x) - E[I(\widehat{u}_t \leq x)|\Omega_t]\} q_{nt}(\widehat{u}_{t-j}, y, \lambda),\end{aligned}$$

where q_{nt} is implicitly defined. Then, to handle A_{1n} , we apply a version of Lemma A2 with $\widehat{c} = \sqrt{n}(\widehat{\theta}_n - \theta_0) = O_p(1)$ and get

$$\sup_{\eta=(\lambda, x, y) \in \Pi} \frac{1}{\sqrt{n}} \left| \sum_{t=2}^n \{I(\widehat{u}_t \leq x) - E[I(\widehat{u}_t \leq x)|\Omega_t]\} q_{nt}(\widehat{u}_{t-j}, y, \lambda) - \sum_{t=2}^n \{I(u_t \leq x) - E[I(u_t \leq x)|\Omega_t]\} q_{nt}(u_{t-j}, y, \lambda) \right| = o_p(1).$$

Hence, uniformly in $\eta = (\lambda, x, y) \in \Pi$,

$$\begin{aligned}A_{1n}(x, y, \lambda) &= \frac{1}{\sqrt{n}} \sum_{t=2}^n \{E[I(\widehat{u}_t \leq x)|\Omega_t] q_{nt}(\widehat{u}_{t-j}, y, \lambda) - E[I(u_t \leq x)|\Omega_t] q_{nt}(u_{t-j}, y, \lambda)\} + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=2}^n \{E[I(\widehat{u}_t \leq x)|\Omega_t] - E[I(u_t \leq x)|\Omega_t]\} q_{nt}(\widehat{u}_{t-j}, y, \lambda) + \\ &\quad \frac{1}{\sqrt{n}} \sum_{t=2}^n E[I(u_t \leq x)|\Omega_t] \{q_{nt}(\widehat{u}_{t-j}, y, \lambda) - q_{nt}(u_{t-j}, y, \lambda)\} + o_p(1) \\ &\equiv B_{1n}(x, y, \lambda) + B_{2n}(x, y, \lambda) + o_p(1).\end{aligned}$$

By the Mean Value Theorem, the ULLN of Newey and McFadden (1994) and Lemma 1 in Escanciano and Velasco (2006), we obtain, in $L_2(\Pi, v)$,

$$\begin{aligned}B_{1n}(x, y, \lambda) &= \sqrt{n}(\widehat{\theta}_n - \theta_0)' \sum_{j=1}^{n-1} \left(\frac{1}{\sqrt{n}} \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \frac{\partial F_t(\widehat{\theta}_n, x)}{\partial \theta} I(\widehat{u}_{t-j} \leq y) \right) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi} \\ &= \sqrt{n}(\widehat{\theta}_n - \theta_0)' F(y) E \left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} \right] \sum_{j=1}^{\infty} \frac{\sqrt{2} \sin j\pi\lambda}{j\pi} + o_p(1).\end{aligned}$$

Similarly, in $L_2(\Pi, \nu)$,

$$B_{2n}(x, y, \lambda) = \sqrt{n}(\hat{\theta}_n - \theta_0)' F(x) E \left[\frac{\partial F_t(\theta_0, y)}{\partial \theta} \right] \sum_{j=1}^{\infty} \frac{\sqrt{2} \sin j\pi\lambda}{j\pi} + o_p(1).$$

In a similar way, we can prove that

$$\|A_{2n} - B_{1n} - B_{2n}\| = o_P(1),$$

and hence, in $L_2(\Pi, \nu)$,

$$\hat{S}_n(\eta) = S_n(\eta) + o_p(1).$$

Finally, notice that $|\hat{T}_{GCM} - \|\hat{S}_n(\eta)\|^2| = o_p(1)$, $|T_{GCM} - \|S_n(\eta)\|^2| = o_p(1)$, and $T_{GCM} \xrightarrow{d} T_\infty = \int_{[0,1]^3} (Z(\eta))^2 d\eta$, (Hong 2000, Theorem 3), and we then get $\hat{T}_{GCM} \xrightarrow{d} T_\infty$.

■

PROOF OF THEOREM 3: Applying the ULLN of Newey and McFadden (1994), Lemma 2.4, to the empirical process

$$\frac{1}{n} \sum_{t=1+j}^n I(u_t(\theta) \leq x) I(u_{t-j}(\theta) \leq y), \quad \theta \in \Theta_0, \quad x \in \mathbb{R}, \quad y \in \mathbb{R},$$

we get

$$\sup_{(\theta, x, y) \in \Theta_0 \times \mathbb{R}^2} \left| \frac{1}{n} \sum_{t=1+j}^n I(u_t(\theta) \leq x) I(u_{t-j}(\theta) \leq y) - E[I(u_t(\theta) \leq x) I(u_{t-j}(\theta) \leq y)] \right| = o_p(1),$$

and hence

$$\sup_{(x, y) \in \mathbb{R}^2} \left| \frac{1}{n} \sum_{t=1+j}^n I(u_t(\hat{\theta}_n) \leq x) I(u_{t-j}(\hat{\theta}_n) \leq y) - E[I(u_t(\hat{\theta}_n) \leq x) I(u_{t-j}(\hat{\theta}_n) \leq y)] \right| = o_p(1).$$

Similarly, we have

$$\sup_{(x, y) \in \mathbb{R}^2} \left| \frac{1}{n} \sum_{t=1}^n I(u_t(\hat{\theta}_n) \leq x) - E[I(u_t(\hat{\theta}_n) \leq x)] \right| = o_p(1).$$

Therefore,

$$\begin{aligned} & \sup_{(x, y) \in \mathbb{R}^2} |\hat{\gamma}_n(x, y) - \gamma_n(x, y)| \\ & \leq \sup_{(x, y) \in \mathbb{R}^2} \left| E[I(u_t(\hat{\theta}_n) \leq x) I(u_{t-j}(\hat{\theta}_n) \leq y)] - E[I(u_t \leq x) I(u_{t-j} \leq y)] \right| \\ & \quad + \sup_{(x, y) \in \mathbb{R}^2} \left| E[I(u_t(\hat{\theta}_n) \leq x)] E[I(u_{t-j}(\hat{\theta}_n) \leq y)] - E[I(u_t \leq x)] E[I(u_{t-j} \leq y)] \right| + o_p(1) \\ & = o_p(1), \end{aligned}$$

where the last inequality follows from the continuity of the mappings $\theta \rightarrow E[I(u_t(\theta) \leq x) I(u_{t-j}(\theta) \leq y)]$ and $\theta \rightarrow E[I(u_t(\theta) \leq x)]$.

■

REFERENCES

- [1] BILLINGSLEY, P. (1968), *Convergence of Probability Measures*, New York: Wiley.
- [2] BLUM, J.R., KIEFER, J. AND ROSENBLATT, M. (1961), "Distribution free tests of independence based on the sample distribution function", *Annals of Mathematical Statistics*, 32, 485-498.
- [3] BOSQ, D. (2000), *Linear Processes in Function Spaces*, Lecture Notes in Statistics 149. Springer.
- [4] BOX, G. AND PIERCE, D. (1970), "Distribution of residual autocorrelations in autoregressive integrated moving average time series models", *Journal of the American Statistical Association* 65, 1509-1527.
- [5] BRAVO, F. (2007), "Two-step generalized empirical likelihood inference for semiparametric models", working paper.
- [6] BROCK, W., DECHERT, D., SCHEINKMAN, J. AND LEBARON, B. (1996), "A test for independence based on the correlation dimension", *Econometric Reviews* 15, 197-235.
- [7] BROCK, W., HSIEH, D. AND LEBARON, B. (1991), *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*, Cambridge: MIT Press.
- [8] CAI, T.X., TIAN, L., SOLOMON, S.D. AND WEI, L.J. (2008), "Predicting future responses based on possibly misspecified working models", *Biometrika* 95, 1, pp. 75-92.
- [9] DELGADO, M.A. (1999), "Hoeffding-Blum-Kiefer-Rosenblat process", *Encyclopedia of Statistical Sciences Update*, 3, Ed. S. Kotz, C.B. Read and D.L. Banks, pp. 326-328. New York: Wiley.
- [10] DELGADO, M.A. AND ESCANCIANO, J.C. (2007), "Nonparametric Tests for Conditional Symmetry in Dynamic Models", *Journal of Econometrics*, 141, 652-682.
- [11] DELGADO, M.A. AND MORA, J. (2000), "A nonparametric test for serial independence of regression errors", *Biometrika* 87, 228-234.
- [12] DURBIN, J. (1973), "Weak convergence of the sample distribution function when parameters are estimated", *Annals of Statistics*, 1, 274-290.
- [13] ESCANCIANO, J.C. (2007), "Weak Convergence of Non-Stationary Multivariate Marked Processes with Applications to Martingale Testing", *Journal of Multivariate Analysis*, 98, 1321-1336.
- [14] ESCANCIANO, J.C. AND OLMO, J. (2010), "Backtesting Parametric Value-at-Risk with Estimation Risk", *Journal of Business and Economics Statistics*, 28, 36-51.

- [15] ESCANCIANO, J.C. AND VELASCO, C. (2006), "Generalized Spectral Tests for the Martingale Difference Hypothesis", *Journal of Econometrics*, 134, 151-185.
- [16] HENDRY, D.F. (1995), *Dynamic Econometrics*, Oxford University Press.
- [17] HOEFDING, W. (1948), "A nonparametric test of independence", *Annals of Mathematical Statistics* 26, 189-211.
- [18] HONG, Y. (2000), "Generalized Spectral Test for Serial Dependence", *Journal of the Royal Statistical Society Series B*, 62, part3. pp. 557-574.
- [19] HONG, Y. AND LEE T.H. (2003), "Diagnostic Checking for Adequacy of Linear and Nonlinear Time Series Models", *Econometric Theory*, 19, 2003, 1065-1121.
- [20] LJUNG, G.M. AND BOX, G. (1978), "On a measure of lack of fit in time series models", *Biometrika* 65, 297-303.
- [21] MCLEOD, A.I. AND LI, W.K. (1983), "Diagnostic checking ARMA time series models using squared-residual autocorrelations", *Journal of Time Series Analysis* 4, 269-273.
- [22] NEWEY, W. K. (1991), "Uniform convergence in probability and stochastic equicontinuity", *Econometrica* 59, 1161-67.
- [23] NEWEY, W.K. AND MCFADDEN, D. (1994), "Large sample estimation and hypothesis testing", *Handbook of Econometrics*, IV, Edited by R.F. Engle and D.L. McFadden, 2111-2244.
- [24] PIERCE, D. (1985), "Testing normality in autoregressive models", *Biometrika*, 72, 293-297.
- [25] VAN DER VAART, A. W. (1998), *Asymptotic Statistics*. Cambridge University Press.
- [26] VAN DER VAART, A. W. AND WELLNER, J.A.(1996), *Weak Convergence and Empirical Processes*. Springer, New York.
- [27] WOODWARD, W.A. AND GRAY, H.L. (1993), "Global warming and the problem of testing for trend in time series data." *Journal of Climate* 6, 953-962.
- [28] WOODWARD, W.A. AND GRAY, H.L. (1995), "Selecting a model for detecting the presence of a trend." *Journal of Climate* 8, 1929-1937.
- [29] XUE, L. AND ZHU, L. (2006), "Empirical likelihood for single index models", *Journal of Multivariate Analysis* 97, 1295-1312.