Statistica Sinica **31** (2021), 269-300
doi:https://doi.org/10.5705/ss.202018.0159

NEW HSIC-BASED TESTS FOR INDEPENDENCE BETWEEN TWO STATIONARY MULTIVARIATE TIME SERIES

Guochang Wang, Wai Keung Li and Ke Zhu

Jinan University, The Education University of Hong Kong and The University of Hong Kong

Abstract: We propose novel one-sided omnibus tests for independence between two multivariate stationary time series. These new tests apply the Hilbert–Schmidt independence criterion (HSIC) to test the independence between the innovations of the time series. We establish the limiting null distributions of our HSIC-based tests under regular conditions. Next, our HSIC-based tests are shown to be consistent. A residual bootstrap method is used to obtain the critical values for the tests, and its validity is justified. Existing cross-correlation-based tests examine linear dependence. In contrast, our tests examine general dependence (including linear and non-linear), providing researchers with information that is more complete on the causal relationship between two multivariate time series. The merits of our tests are illustrated using simulations and a real-data example.

Key words and phrases: Hilbert-Schmidt independence criterion, multivariate time series models, non-linear dependence, residual bootstrap, testing for independence.

1. Introduction

Before applying a sophisticated method to describe the relationship between two time series, it is important to first determine whether they are independent. If the dependence exists, causal analysis techniques, such as copula and multivariate modeling, can be used to investigate the relationship between them, potentially leading to interesting insights or effective predictive models. However, if two time series are independent, one should use two independent parsimonious models; see, for example, Pierce (1977); Schwert (1979); Hong (2001a); Lee and Long (2009); Shao (2009); Tchahou and Duchesne (2013) for empirical examples in this context.

Most existing methods used to test independence between two multivariate time series models apply a measure based on cross-correlations. Specifically,

Corresponding author: Ke Zhu, Department of Statistics & Actuarial Science, The University of Hong Kong, Hong Kong. E-mail: mazhuke@hku.hk.

they check whether the sample cross-correlations of the model residuals are significantly different from zero, up to a fixed lag or all valid lags. The former includes the portmanteau tests (Cheung and Ng (1996); El Himdi and Roy (1997); Pham, Roy and Cédras (2003); Hallin and Saidi (2005, 2007); Robbins and Fisher (2015)), and the latter (with the aid of kernel smoothing) is a type of spectral test (Hong (2001a,b); Bouhaddioui and Roy (2006)). Note that the idea of using cross-correlations in this way is a natural extension of the seminal studies of Haugh (1976) and Hong (1996) for univariate time series models. However, in many circumstances, this conveys uncorrelatedness, rather than independence.

In general, the aforementioned tests are designed to investigate linear dependence (i.e., the cross-correlation in the mean, variance, or higher moments) between two model residuals, and hence may lack power in detecting a non-linear dependence structure. A significant body of research has documented non-linear dependence relationships between various economic fundamentals; see, for example, Hiemstra and Jones (1994); Wang, Wu and Yang (2013); Choudhry, Papadimitriou and Shabi (2016); Diks and Wolski (2016), among others. However, few studies attempt to account for both linear and non-linear dependence, both of which are important characteristics.

To examine the general dependence structure, a test needs a direct measure of independence. In the last decade, the Hilbert–Schmidt independence criterion (HSIC) of Gretton et al. (2005) has been used extensively in many fields. Works that provide one- or two-sample independence tests based on the HSIC include those of Gretton et al. (2008) and Gretton and Györfi (2010) for observable independent and identically distributed (i.i.d.) data, and Zhang et al. (2009); Zhou (2012); Fokianos and Pitsillou (2017) for observable dependent or time series data. The latter two studies applied the distance covariance (DC) of Székely, Rizzoand Bakirov (2007), whereas Sejdinovic et al. (2013) showed that the HSIC and DC are equivalent. When the data are unobservable and are derived from a fitted statistical model (e.g., the estimated model innovations), the estimation effect has to be considered. The original procedure based on the HSIC or DC is no longer valid; thus, we need to modify the procedure for testing purposes. However, very little work has been done in this context. Two exceptions are Sen and Sen (2014) and Davis et al. (2018) for one-sample independence tests. The former focused on a regression model with independent covariates, and the latter considered vector AR models, but without providing a rigorous way to obtain the critical values of the related test.

This paper proposes novel one-sided tests for the independence of two stationary multivariate time series. These new tests apply the HSIC to examine the independence between the unobservable innovation vectors of both time series. Of these tests, the single HSIC-based test is tailored to detect general dependence between these two innovation vectors at a specific lag m, and the joint HSIC-based test is designed for this purpose up to certain lag M. Under regular conditions, the limiting null distributions of our HSIC-based tests are established. Next, our HSIC-based tests are shown to be consistent. Moreover, a residual bootstrap method is used to obtain the critical values for our tests, and its validity is justified. Our methodologies are applicable for general specifications of time series models driven by i.i.d. innovations. By choosing different lags, our tests provide investigators with information that is more complete on the general dependence (including both linear and non-linear) relationship between two time series. Finally, the importance of our HSIC-based tests is illustrated using simulations and a real-data example.

This paper is organized as follows. Section 2 introduces our HSIC-based test statistics. Section 3 studies the asymptotic properties of our HSIC-based tests. A residual bootstrap method is provided in Section 4. Simulation results are reported in Section 5. A real-data example is presented in Section 6, and concluding remarks are offered in Section 7. Additional simulations and the proofs are provided in the online Supplementary Material.

Throughout the paper, $\mathcal{R} = (-\infty, \infty)$, C is a generic constant, I_s is the $s \times s$ identity matrix, 1_s is the $s \times 1$ vector of ones, $\otimes$ is the Kronecker product, A^T is the transpose of matrix A, $\|A\|$ is the Euclidean norm of matrix A, $vec(A)$ is the vectorization of A, $vech(A)$ is the half vectorization of A, $D(A)$ is the diagonal matrix whose main diagonal is the main diagonal of matrix A, $\partial_x h$ denotes the partial derivative with respect to x, for any function $h(x, y, \ldots)$, $o_p(1)(O_p(1))$ denotes a sequence of random numbers converging to zero (bounded) in probability, "$\rightarrow_d$" denotes convergence in distribution, and "$\rightarrow_p$" denotes convergence in probability.

2. The HSIC-based Test Statistics

2.1. Review of the HSIC

In this subsection, we briefly review the HSIC, which tests the independence of two random vectors; see, for example, Gretton et al. (2005) and Gretton et al. (2008) for more detail.

Let $\mathcal{U}$ be a metric space, and let $k : \mathcal{U} \times \mathcal{U} \rightarrow \mathcal{R}$ be a symmetric and positivedefinite (i.e., $\sum_{i,j} c_i c_j k(x_i, x_j) \geq 0$ for all $c_i \in \mathcal{R}$) kernel function. There exists a Hilbert space $\mathcal{H}$ (called a *Reproducing Kernel Hilbert Space* (RKHS)) of functions $f : \mathcal{U} \to \mathcal{R}$ with inner product $\langle \cdot, \cdot \rangle$, such that

$$(i)\ \ k(u, \cdot) \in \mathcal{H},\ \text{ for } \forall u \in \mathcal{U}; \tag{2.1}$$

$$(ii)\ \ \langle f, k(u, \cdot) \rangle = f(u),\ \text{ for } \forall f \in \mathcal{H} \text{ and } \forall u \in \mathcal{U}. \tag{2.2}$$

For any Borel probability measure P defined on $\mathcal{U}$, its *mean element* $\mu[P] \in \mathcal{H}$ is defined as follows:

$$E[f(U)] = \langle f, \mu[P] \rangle, \quad \forall f \in \mathcal{H}, \tag{2.3}$$

where the random variable $U \sim P$. From (2.2)–(2.3), we have $\mu[P](u) = \langle k(\cdot, u), \mu[P] \rangle = E[k(U, u)]$. Furthermore, we say that $\mathcal{H}$ is *characteristic* if and only if the map $P \to \mu[P]$ is injective on the space $\mathcal{P} := \{P : \int_{\mathcal{U}} k(u, u) dP(u) < \infty\}$.

Likewise, let $\mathcal{G}$ be a second RKHS on a metric space $\mathcal{V}$ with kernel l. Let P_{uv} be a Borel probability measure defined on $\mathcal{U} \times \mathcal{V}$, and let P_u and P_v denote the marginal distributions on $\mathcal{U}$ and $\mathcal{V}$, respectively. Assume that

$$E[k(U, U)] < \infty \quad \text{and} \quad E[l(V, V)] < \infty, \tag{2.4}$$

where the random variable $(U, V) \sim P_{uv}$. The HSIC of P_{uv} is defined as

$$\begin{aligned}\Pi(U, V) :=\ & E_{U,V} E_{U',V'} [k(U, U') l(V, V')] + E_U E_{U'} E_V E_{V'} [k(U, U') l(V, V')] \\ & -2 E_{U,V} E_{U'} E_{V'} [k(U, U') l(V, V')],\end{aligned}$$

where (U', V') is an i.i.d. copy of (U, V), and $E_{\xi,\zeta}$ (or E_ξ) denotes the expectation over (ξ, ζ) (or ξ). Following Sejdinovic et al. (2013), if (2.4) holds and both $\mathcal{H}$ and $\mathcal{G}$ are characteristic, then

$$\Pi(U, V) = 0 \quad \text{if and only if} \quad P_{uv} = P_u \times P_v.$$

Therefore, we can test the independence of U and V by examining whether $\Pi(U, V)$ is significantly different from zero.

Suppose the samples $\{(U_i, V_i)\}_{i=1}^n$ are from P_{uv}. Following Gretton et al. (2005), the empirical estimator of $\Pi(U, V)$ is

$$\Pi_n = \frac{1}{n^2} \sum_{i,j} k_{ij} l_{ij} + \frac{1}{n^4} \sum_{i,j,q,r} k_{ij} l_{qr} - \frac{2}{n^3} \sum_{i,j,q} k_{ij} l_{iq} \tag{2.5}$$

$$= \frac{1}{n^2}\text{trace}(KHLH), \tag{2.6}$$

where $k_{ij} = k(U_i, U_j)$; $l_{ij} = l(V_i, V_j)$; $K = (k_{ij})$ and $L = (l_{ij})$ are $n \times n$ matrices, with entries k_{ij} and l_{ij}, respectively; and $H = I_n - (1_n 1_n^T)/n$. Here, each index of the summation $\sum$ is taken from 1 to n. If $\{(U_i, V_i)\}_{i=1}^n$ are i.i.d. samples, Gretton et al. (2005) showed that Π_n is a consistent estimator of $\Pi(U, V)$.

In order to compute Π_n, we need to choose the kernel functions k and l. In what follows, we assume $\mathcal{U} = \mathcal{R}^{\kappa_1}$ and $\mathcal{V} = \mathcal{R}^{\kappa_2}$, for two positive integers κ_1 and κ_2. Then, the following are some well-known choices (see Peters (2008); Zhang et al. (2018)) for k (or l):

$$\begin{aligned}
\text{[Gaussian kernel]} : k(u, u') &= \exp\left(-\frac{\|u - u'\|^2}{2\sigma^2}\right), \\
&\quad \text{for some } \sigma > 0; \\
\text{[Laplace kernel]} : k(u, u') &= \exp\left(-\frac{\|u - u'\|}{\sigma}\right), \\
&\quad \text{for some } \sigma > 0; \\
\text{[Inverse multi-quadratics kernel]} : k(u, u') &= \frac{1}{(\beta + \|u - u'\|)^\alpha}, \\
&\quad \text{for some } \alpha, \beta > 0; \\
\text{[Fractional Brownian motion kernel]} : k(u, u') &= \frac{1}{2}(\|u\|^{2h} + \|u'\|^{2h} - \|u - u'\|^{2h}), \\
&\quad \text{for some } 0 < h < 1.
\end{aligned}$$

Note that the HSIC is easy to implement in multivariate cases, because the computation cost of Π_n is $O(n^2)$, regardless of the dimensions of U and V, and many software packages can calculate (2.6) very quickly.

2.2. Test statistics

Consider two multivariate time series Y_{1t} and Y_{2t}, where $Y_{1t} \in \mathcal{R}^{d_1}$ and $Y_{2t} \in \mathcal{R}^{d_2}$. Assume that each Y_{st} ($s = 1$ or 2, hereafter) admits the following specification:

$$Y_{st} = f_s(I_{st-1}, \theta_{s0}, \eta_{st}), \tag{2.7}$$

where $I_{st} = (Y_{st}^T, Y_{st-1}^T, \ldots)^T \in \mathcal{R}^\infty$ is the information set at time t; $\theta_{s0} \in \mathcal{R}^{p_s}$ is the true, but unknown parameter value of model (2.7); $\eta_{st} \in \mathcal{R}^{d_s}$ is a sequence of i.i.d. innovations, such that η_{st} and $\mathcal{F}_{st-1}$ are independent; $\mathcal{F}_{st} := \sigma(I_{st})$ is a sigma-field; and $f_s : \mathcal{R}^\infty \times \mathcal{R}^{p_s} \times \mathcal{R}^{d_s} \to \mathcal{R}^{d_s}$ is a known measurable function. Model (2.7) is rich enough to include many often-used models, such as the vector

AR model of Sims (1980), BEKK model of Engle and Kroner (1995), dynamic correlation model of Tse (2002), and vector ARMA-GARCH model of Ling and McAleer (2003), among others; see also Lütkepohl (2005); Bauwens, Laurent and Rombouts (2006); Silvennoinen and Teräsvirta (2009); Francq and Zakoïan (2010); Tsay (2014) for surveys.

Model (2.7) ensures that each Y_{st} admits a dynamical system generated by the innovation sequence $\{\eta_{st}\}$. A practical question is whether either one of the dynamical systems should include information on the other, which is equivalent to testing the following null hypothesis:

$$H_0 : \{\eta_{1t}\} \text{ and } \{\eta_{2t}\} \text{ are independent.} \tag{2.8}$$

If H_0 is accepted, we can separately study these two systems; otherwise, we may use the information of one system to obtain a better prediction of the other system. Let m be a given integer. Most conventional testing methods for H_0 in (2.8) aim to detect linear dependence between η_{1t} and η_{2t+m} (or their higher moments) using their cross-correlations. Below, we apply the HSIC to examine the general dependence between η_{1t} and η_{2t+m}.

To introduce our HSIC-based tests, we need some additional notation. Let $\theta_s = (\theta_{s1}, \theta_{s2}, \ldots, \theta_{sp_s}) \in \Theta_s \subset \mathcal{R}^{p_s}$ be the unknown parameter of model (2.7), where Θ_s is a compact parametric space. Assume that θ_{s0} is an interior point of Θ_s, and Y_{st} admits a causal representation:

$$\eta_{st} = g_s(Y_{st}, I_{st-1}, \theta_{s0}), \tag{2.9}$$

where $g_s : \mathcal{R}^{d_s} \times \mathcal{R}^{\infty} \times \mathcal{R}^{p_s} \to \mathcal{R}^{d_s}$ is a measurable function. Moreover, based on the observations $\{Y_{st}\}_{t=1}^{n}$ and (possibly) some assumed initial values, we let

$$\widehat{\eta}_{st} := g_s(Y_{st}, \widehat{I}_{st-1}, \widehat{\theta}_{sn}) \tag{2.10}$$

be the residual of model (2.7), where $\widehat{\theta}_{sn}$ is an estimator of θ_{s0}, and $\widehat{I}_{st}$ is the observed information set up to time t.

As in (2.5)–(2.6), our single HSIC-based test statistic on $\widehat{\eta}_{1t}$ and $\widehat{\eta}_{2t+m}$ is

$$\begin{aligned} S_{1n}(m) := \Pi(\widehat{\eta}_{1t}, \widehat{\eta}_{2t+m}) &= \frac{1}{N^2}\sum_{i,j}\widehat{k}_{ij}\widehat{l}_{ij} + \frac{1}{N^4}\sum_{i,j,q,r}\widehat{k}_{ij}\widehat{l}_{qr} - \frac{2}{N^3}\sum_{i,j,q}\widehat{k}_{ij}\widehat{l}_{iq} \\ &= \frac{1}{N^2}\mathrm{trace}(\widehat{K}H\widehat{L}H), \end{aligned} \tag{2.11}$$

for $m \geq 0$, where $\widehat{k}_{ij} = k(\widehat{\eta}_{1i}, \widehat{\eta}_{1j})$, $\widehat{l}_{ij} = l(\widehat{\eta}_{2i+m}, \widehat{\eta}_{2j+m})$, and $\widehat{K} = (\widehat{k}_{ij})$ and $\widehat{L} = (\widehat{l}_{ij})$ are $N \times N$ matrices with entries $\widehat{k}_{ij}$ and $\widehat{l}_{ij}$, respectively. Here, the effective sample size $N = n - m$, and each index of the summation is taken from 1 to N. Likewise, our single HSIC-based test statistic on $\widehat{\eta}_{1t+m}$ and $\widehat{\eta}_{2t}$ is

$$S_{2n}(m) := \Pi(\widehat{\eta}_{1t+m}, \widehat{\eta}_{2t}), \tag{2.12}$$

for $m \geq 0$. Clearly, $S_{1n}(0) = S_{2n}(0)$.

Using the single HSIC-based test statistics, we can further define the joint HSIC-based test statistics as follows:

$$J_{1n}(M) := \sum_{m=0}^{M} S_{1n}(m) \quad \text{and} \quad J_{2n}(M) := \sum_{m=0}^{M} S_{2n}(m), \tag{2.13}$$

for some specified integer $M \geq 0$. The joint test statistic, $J_{1n}(M)$ or $J_{2n}(M)$, can detect the general dependence structure of two innovations up to certain lag M; in contrast, the single test statistic, $S_{1n}(m)$ or $S_{2n}(m)$, is used to examine the general dependence structure of two innovations at a specific lag m.

3. Asymptotic Theory

This section studies the asymptotics of our HSIC-based test statistics $S_{1n}(m)$ and $J_{1n}(M)$. The asymptotics of $S_{2n}(m)$ and $J_{2n}(M)$ can be derived similarly, and hence the details are omitted for simplicity.

3.1. Technical conditions

To derive our asymptotic theory, the following assumptions are needed.

Assumption 1. *Y_{st} is strictly stationary and ergodic.*

Assumption 2. *(i) The function $g_{st}(\theta_s) := g_s(Y_{st}, I_{st-1}, \theta_s)$ satisfies that*

$$E\left[\sup_{\theta_s}\left\|\frac{\partial g_{st}(\theta_s)}{\partial \theta_{si}}\right\|\right]^2 < \infty, \quad E\left[\sup_{\theta_s}\left\|\frac{\partial^2 g_{st}(\theta_s)}{\partial \theta_{si}\partial \theta_{sj}}\right\|\right]^2 < \infty,$$
$$\text{and} \quad E\left[\sup_{\theta_s}\left\|\frac{\partial^3 g_{st}(\theta_s)}{\partial \theta_{si}\partial \theta_{sj}\partial \theta_{sq}}\right\|\right]^2 < \infty,$$

for any i, j, $q \in \{1, \ldots, p_s\}$, where g_s is defined as in (2.9).

(ii) $\sum_{j=0}^{\infty} \beta_\eta(j)^{c/(2+c)} < \infty$, for some $c > 0$, where $\beta_\eta(j)$ is the β-mixing coefficient of $\{(\eta_{1t}^T, \eta_{2t}^T)^T\}$.

Assumption 3. *The estimator* $\widehat{\theta}_{sn}$ *given in* (2.10) *satisfies that*

$$\sqrt{n}(\widehat{\theta}_{sn} - \theta_{s0}) = \frac{1}{\sqrt{n}} \sum_t \pi_s(Y_{st}, I_{st-1}, \theta_{s0}) + o_p(1)$$
$$=: \frac{1}{\sqrt{n}} \sum_t \pi_{st} + o_p(1), \quad (3.1)$$

where $\pi_s : \mathcal{R}^{d_s} \times \mathcal{R}^\infty \times \mathcal{R}^{p_s} \to \mathcal{R}^{p_s}$ *is a measurable function,* $E(\pi_{st}|\mathcal{F}_{st-1}) = 0$, *and* $E\|\pi_{st}\|^2 < \infty$.

Assumption 4. *For* $\widehat{R}_{st}(\theta_s) := \widehat{g}_{st}(\theta_s) - g_{st}(\theta_s)$,

$$\sum_t \sup_{\theta_s} \|\widehat{R}_{st}(\theta_s)\|^3 = O_p(1),$$

where $\widehat{g}_{st}(\theta_s) = g_s(Y_{st}, \widehat{I}_{st-1}, \theta_s)$, *and* $\widehat{I}_{st}$ *is defined as in* (2.10).

Assumption 5. *The kernel functions* k *and* l *are symmetric. Furthermore, both of them and their partial derivatives up to the second order are uniformly bounded and Lipschitz continuous, that is,*

$$(i)\ \sup_{x,y} \|p(x,y)\| \le C; \ (ii)\ \|p(x_1,y_1) - p(x_2,y_2)\| \le C\|(x_1,y_1) - (x_2,y_2)\|,$$

for $p = k, k_x, k_y, k_{xx}, k_{xy}, k_{yy}, l, l_x, l_y, l_{xx}, l_{xy}, l_{xy}$, *where* $k_x = \partial_x k(x,y)$, $k_{xy} = \partial_x \partial_y k(x,y)$, $l_x = \partial_x l(x,y)$, *and* $l_{xy} = \partial_x \partial_y l(x,y)$.

A few remarks are in order related to the above assumptions. Assumption 1 is standard for time series models. Assumption 2(i) requires technical moment conditions for the partial derivatives of g_{st}. Assumption 2(ii) gives a sufficient technical condition to prove Theorem 2, for which the result of part (c) of Theorem 1 in Denker and Keller (1983) can be applied directly. Assumption 3 is satisfied under mild conditions for most estimators, including the (quasi) maximum likelihood estimator (MLE), least squares estimator (LSE), nonlinear least squares estimator (NLSE), and their robust modifications; see, for example, Comte and Lieberman (2003); Lütkepohl (2005) and Hafner and Preminger (2009) for further detail. Assumption 4 is a condition on the truncation of the information set $\widehat{I}_{st-1}$, and is similar to Assumption A5 in Escanciano (2006). Assumption 5 provides restrictive conditions for the kernel functions k and l. These conditions may exclude some kernel functions, such as the fractional Brownian motion kernel, but they are usually satisfied by the often-used Gaussian kernel, Laplace kernel, and inverse multi-quadratics kernel. The conditions in Assumptions 1–5 may be relaxed further, but they are convenient for presenting our proofs in a simple way.

3.2. Lemmas

This subsection provides several lemmas, that are important to derive the asymptotics of our test statistics.

Before introducing these lemmas, we present some additional notation. Let

$$\overline{k}_{ij} = \frac{\partial g_{1i}(\theta_{10})}{\partial \theta_1} k_x(\eta_{1i}, \eta_{1j}) + \frac{\partial g_{1j}(\theta_{10})}{\partial \theta_1} k_y(\eta_{1i}, \eta_{1j}), \tag{3.2}$$

$$\bar{l}_{qr} = \frac{\partial g_{2q+m}(\theta_{20})}{\partial \theta_2} l_x(\eta_{2q+m}, \eta_{2r+m}) + \frac{\partial g_{2r+m}(\theta_{20})}{\partial \theta_2} l_y(\eta_{2q+m}, \eta_{2r+m}), \tag{3.3}$$

$$\begin{aligned} \breve{k}_{ij} = & \left(\frac{\partial g_{1i}(\theta_{10})}{\partial \theta_1}, \frac{\partial g_{1j}(\theta_{10})}{\partial \theta_1} \right) \begin{pmatrix} k_{xx}(\eta_{1i}, \eta_{1j}) & k_{xy}(\eta_{1i}, \eta_{1j}) \\ k_{xy}(\eta_{1i}, \eta_{1j}) & k_{yy}(\eta_{1i}, \eta_{1j}) \end{pmatrix} \\ & \times \left(\frac{\partial g_{1i}(\theta_{10})}{\partial \theta_1}, \frac{\partial g_{1j}(\theta_{10})}{\partial \theta_1} \right)^T, \end{aligned} \tag{3.4}$$

$$\begin{aligned} \breve{l}_{qr} = & \left(\frac{\partial g_{2q+m}(\theta_{20})}{\partial \theta_2}, \frac{\partial g_{2r+m}(\theta_{20})}{\partial \theta_2} \right) \\ & \times \begin{pmatrix} l_{xx}(\eta_{2q+m}, \eta_{2r+m}) & l_{xy}(\eta_{2q+m}, \eta_{2r+m}) \\ l_{xy}(\eta_{2q+m}, \eta_{2r+m}) & l_{yy}(\eta_{2q+m}, \eta_{2r+m}) \end{pmatrix} \\ & \times \left(\frac{\partial g_{2q+m}(\theta_{20})}{\partial \theta_2}, \frac{\partial g_{2r+m}(\theta_{20})}{\partial \theta_2} \right)^T, \end{aligned} \tag{3.5}$$

for $i, j, q, r \in \{1, 2, \ldots, N\}$. With these notation, define

$$S_{1n}^{(0)}(m) = \frac{1}{N^2} \sum_{i,j} k_{ij} l_{ij} + \frac{1}{N^4} \sum_{i,j,q,r} k_{ij} l_{qr} - \frac{2}{N^3} \sum_{i,j,q} k_{ij} l_{iq}, \tag{3.6}$$

$$S_{1n}^{(ab)}(m) = \frac{1}{N^2} \sum_{i,j} k_{ij}^{(ab)} l_{ij}^{(ab)} + \frac{1}{N^4} \sum_{i,j,q,r} k_{ij}^{(ab)} l_{qr}^{(ab)} - \frac{2}{N^3} \sum_{i,j,q} k_{ij}^{(ab)} l_{iq}^{(ab)}, \tag{3.7}$$

for $a \in \{1, 2\}$ and $b \in \{1, \ldots, a+1\}$, where $k_{ij}^{(11)} = \overline{k}_{ij}$, $l_{ij}^{(11)} = l_{ij}$, $k_{ij}^{(12)} = k_{ij}$, $l_{ij}^{(12)} = \bar{l}_{ij}$, $k_{ij}^{(21)} = \breve{k}_{ij}$, $l_{ij}^{(21)} = l_{ij}$, $k_{ij}^{(22)} = k_{ij}$, $l_{ij}^{(22)} = \breve{l}_{ij}$, $k_{ij}^{(23)} = \overline{k}_{ij}$, and $l_{ij}^{(23)} = \bar{l}_{ij}^T$. Then, $S_{1n}^{(0)}(m)$ can be expressed as a V-statistic of the form (see

Gretton et al. (2005)):

$$S_{1n}^{(0)}(m) = \frac{1}{N^4} \sum_{i,j,q,r} h_m^{(0)}(\eta_i^{(m)}, \eta_j^{(m)}, \eta_q^{(m)}, \eta_r^{(m)}),$$

for some symmetric kernel $h_m^{(0)}$, given by

$$h_m^{(0)}(\eta_i^{(m)}, \eta_j^{(m)}, \eta_q^{(m)}, \eta_r^{(m)}) = \frac{1}{4!} \sum_{(t,u,v,w)}^{(i,j,q,r)} (k_{tu}l_{tu} + k_{tu}l_{vw} - 2k_{tu}l_{tv}),$$

where the sum is taken over all 4! permutations of (i, j, q, r), and $\eta_t^{(m)} = (\eta_{1t}, \eta_{2t+m}) \in \mathcal{R}^{d_1} \times \mathcal{R}^{d_2}$. Likewise, all $S_{1n}^{(ab)}(m)$ can be expressed as V-statistics for the symmetric kernel $h_m^{(ab)}$, given by

$$h_m^{(ab)}(\varsigma_i^{(m)}, \varsigma_j^{(m)}, \varsigma_q^{(m)}, \varsigma_r^{(m)}) = \frac{1}{4!} \sum_{(t,u,v,w)}^{(i,j,q,r)} \left(k_{tu}^{(ab)}l_{tu}^{(ab)} + k_{tu}^{(ab)}l_{vw}^{(ab)} - 2k_{tu}^{(ab)}l_{tv}^{(ab)}\right),$$

where the sum is taken over all 4! permutations of (i, j, q, r), and

$$\varsigma_t^{(m)} = \left(\eta_{1t}, \frac{\partial g_{1t}(\theta_{10})}{\partial \theta_1}, \eta_{2t+m}, \frac{\partial g_{2t+m}(\theta_{20})}{\partial \theta_2}\right) \in \mathcal{R}^{d_1} \times \mathcal{R}^{p_1 \times d_1} \times \mathcal{R}^{d_2} \times \mathcal{R}^{p_2 \times d_2}.$$

Now, we are ready to introduce three lemmas. The first lemma gives an important expansion of $S_{1n}(m)$.

Lemma 1. *$S_{1n}(m)$ admits the following expansion:*

$$\begin{aligned} S_{1n}(m) = & S_{1n}^{(0)}(m) + \zeta_{1n}^T S_{1n}^{(11)}(m) + \zeta_{2n}^T S_{1n}^{(12)}(m) \\ & + \frac{1}{2}\zeta_{1n}^T S_{1n}^{(21)}(m)\zeta_{1n} + \frac{1}{2}\zeta_{2n}^T S_{1n}^{(22)}(m)\zeta_{2n} + \zeta_{1n}^T S_{1n}^{(23)}(m)\zeta_{2n} + R_{1n}(m), \end{aligned}$$

where $S_{1n}^{(0)}(m)$ and $S_{1n}^{(ab)}(m)$ are defined as in (3.6) *and* (3.7), *respectively, $R_{1n}(m)$ is the remainder term, and $\zeta_{sn} = \widehat{\theta}_{sn} - \theta_{s0}$.*

The second lemma is crucial to derive the asymptotics of $S_{1n}^{(0)}(m)$ and $S_{1n}^{(ab)}(m)$ under H_0.

Lemma 2. *Suppose Assumptions* 1, 2*(i), and* 5 *hold. Then, under H_0,*

$$(i) \quad E\left[h_m^{(0)}(x_1, \eta_2^{(m)}, \eta_3^{(m)}, \eta_4^{(m)})\right] = 0,$$

for all $x_1 \in \mathcal{R}^{d_1} \times \mathcal{R}^{d_2}$*;*

$$(ii) \quad E\left[h_m^{(ab)}(x_1, \varsigma_2^{(m)}, \varsigma_3^{(m)}, \varsigma_4^{(m)})\right] = 0,$$

for all $x_1 \in \mathcal{R}^{d_1} \times \mathcal{R}^{p_1 \times d_1} \times \mathcal{R}^{d_2} \times \mathcal{R}^{p_2 \times d_2}$ *and each* $a, b = 1, 2$*;*

$$(iii) \quad E\left[h_m^{(23)}(x_1, \varsigma_2^{(m)}, \varsigma_3^{(m)}, \varsigma_4^{(m)})\right] = \Upsilon,$$

for all $x_1 \in \mathcal{R}^{d_1} \times \mathcal{R}^{p_1 \times d_1} \times \mathcal{R}^{d_2} \times \mathcal{R}^{p_2 \times d_2}$*, where*

$$\Upsilon = 4E\left[\frac{\partial g_{12}(\theta_{10})}{\partial \theta_1} k_x(\eta_{12}, \eta_{11})\right] E\left[\frac{\partial g_{22}(\theta_{20})}{\partial \theta_2} l_x(\eta_{22}, \eta_{21}) - \frac{\partial g_{23}(\theta_{20})}{\partial \theta_2} l_x(\eta_{23}, \eta_{21})\right]$$
$$+4E\left[\frac{\partial g_{13}(\theta_{10})}{\partial \theta_1} k_x(\eta_{13}, \eta_{11})\right] E\left[\frac{\partial g_{23}(\theta_{20})}{\partial \theta_2} l_x(\eta_{23}, \eta_{21}) - \frac{\partial g_{22}(\theta_{20})}{\partial \theta_2} l_x(\eta_{22}, \eta_{21})\right].$$

By standard arguments for V-statistics (see, e.g., Lee (1990)), we have $N[S_{1n}^{(0)}(m)] = N[V_{1n}^{(0)}(m)] + o_p(1)$, where

$$V_{1n}^{(0)}(m) = \frac{1}{N^2} \sum_{i,j} h_{2m}^{(0)}(\eta_i^{(m)}, \eta_j^{(m)}) \tag{3.8}$$

is the V-statistic with the kernel function

$$h_{2m}^{(0)}(x_1, x_2) = E\left[h_m^{(0)}(x_1, x_2, \eta_3^{(m)}, \eta_4^{(m)})\right], \tag{3.9}$$

for $x_1, x_2 \in \mathcal{R}^{d_1} \times \mathcal{R}^{d_2}$. Under H_0, $\{\eta_t^{(m)}\}$ is a sequence of i.i.d. random variables. Hence, Lemma 2(i) implies that $V_{1n}^{(0)}(m)$ is a degenerate V-statistic of order 1, from which $h_{2m}^{(0)}$ can be expressed as

$$h_{2m}^{(0)}(x_1, x_2) = \sum_{j=0}^{\infty} \lambda_{jm} \Phi_{jm}(x_1) \Phi_{jm}(x_2), \tag{3.10}$$

where $\{\Phi_{jm}(\cdot)\}$ is an orthonormal function in the L_2-norm, and λ_{jm} is the eigenvalue corresponding to the eigenfunction $\Phi_{jm}(\cdot)$. That is, $\{\lambda_{jm}\}$ is a finite enumeration of the nonzero eigenvalues of the equation

$$E\left[h_{2m}^{(0)}(x_1, \eta_1^{(m)}) \Phi_{jm}(\eta_1^{(m)})\right] = \lambda_{jm} \Phi_{jm}(x_1),$$

where $E\Phi_{jm}(\eta_1^{(m)})=0$ for all $j\geq 1$, and

$$E\left[\Phi_{jm}(\eta_1^{(m)})\Phi_{j'm}(\eta_1^{(m)})\right]=\begin{cases}1, & j=j',\\ 0, & j\neq j'\end{cases}$$

(see, e.g., Dunford and Schwartz (1963, p.1087)). From (3.8) and (3.10), under H_0, we have

$$N[S_{1n}^{(0)}(m)]=\sum_{j=1}^{\infty}\lambda_{jm}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Phi_{jm}(\eta_i^{(m)})\right]^2+o_p(1). \tag{3.11}$$

Next, we consider $S_{1n}^{(ab)}(m)$, which results from the estimation effect. Under H_0, $S_{1n}^{(ab)}(m)$ (for $a,b=1,2$) is a degenerate V-statistic of order 1 by Lemma 2(ii). Hence, $N[S_{1n}^{(ab)}(m)]=O_p(1)$, and its related estimation effect is thus negligible, given that $\zeta_{sn}^T N[S_{1n}^{(ab)}(m)]=o_p(1)$. However, under H_0, the estimation effect related to $S_{1n}^{(23)}(m)$ is negligible only when $\Upsilon=0$. This is because when $\Upsilon\neq 0$, $S_{1n}^{(23)}(m)=O_p(1)$ by the law of large numbers for V-statistics. Thus, its related estimation effect is not negligible in this case, based on the ground that $N[\zeta_{1n}^T S_{1n}^{(23)}(m)\zeta_{2n}]=O_p(1)$.

Our third lemma provides a useful central limit theorem.

Lemma 3. *Suppose Assumptions* 1, 2*(i), and* 3–5 *hold. Then, under* H_0*,*

$$\mathcal{T}_n:=\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathcal{T}_{1i}^T,\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{T}_{2i}^T\right)^T\to_d\mathcal{T}:=((\mathcal{Z}_{jm})_{j\geq 1,0\leq m\leq M},(\mathcal{W}_s^T)_{1\leq s\leq 2})^T$$

as $n\to\infty$*, where* $\mathcal{T}_{1i}=((\Phi_{jm}(\eta_i^{(m)}))_{j\geq 1,0\leq m\leq M})^T$*;* $\mathcal{T}_{2i}=\left(\pi_{1i}^T,\pi_{2i}^T\right)^T$*;* $\mathcal{T}$ *is a multivariate normal distribution with mean zero and covariance matrix* $\overline{\mathcal{T}}=E(\mathcal{T}_1\mathcal{T}_1^T)$*, with* $\mathcal{T}_i=(\mathcal{T}_{1i}^T,\mathcal{T}_{2i}^T)^T$*;* $\{\mathcal{Z}_{jm}\}_{j\geq 1}$ *is a sequence of i.i.d.* $N(0,1)$ *random variables; and* $\mathcal{W}_s$ *is a* p_s*-variate normal random variable.*

3.3. Asymptotics of test statistics

Based on Lemmas 1–3, this subsection studies the asymptotics of our test statistics. Let

$$\Lambda_m^{(23)}:=E\left[h_m^{(23)}(\varsigma_1^{(m)},\varsigma_2^{(m)},\varsigma_3^{(m)},\varsigma_4^{(m)})\right]. \tag{3.12}$$

First, we give the limiting null distributions of $S_{1n}(m)$ and $J_{1n}(M)$.

Theorem 1. *Suppose Assumptions* 1, 2*(i), and* 3–5 *hold. Then, under* H_0*,*

$$(i)\quad n[S_{1n}(m)] \to_d \chi_m \quad \textit{for } 0 \le m \le M;$$
$$(ii)\quad n[J_{1n}(M)] \to_d \sum_{m=0}^{M} \chi_m,$$

as $n \to \infty$*, where* χ_m *is defined by*

$$\chi_m = \sum_{j=1}^{\infty} \lambda_{jm} \mathcal{Z}_{jm}^2 + \mathcal{W}_1^T \Lambda_m^{(23)} \mathcal{W}_2.$$

Here, λ_{jm} *is defined as in* (3.10)*, and* $\mathcal{Z}_{jm}$ *and* $\mathcal{W}_s$ *are defined as in Lemma* 3*.*

Theorem 1 shows that $S_{1n}(m)$ and $J_{1n}(M)$ have convergence rate n^{-1} under H_0. Based on this theorem, we reject H_0 at the significance level α if

$$n[S_{1n}(m)] > c_{m\alpha} \quad \text{or} \quad n[J_{1n}(M)] > c_\alpha,$$

where $c_{m\alpha}$ and c_α are the αth upper quantiles of χ_m and $\sum_{m=0}^{M} \chi_m$, respectively. Because the distribution of χ_m depends on $\{Y_{st}\}$ and $\{\pi_{st}\}$, a residual bootstrap method is proposed in Section 4 to obtain the values of $c_{m\alpha}$ and c_α.

Second, we study the behavior of $S_{1n}(m)$ under the following fixed alternative:

$$H_1^{(m)}: \ \{\eta_{1t}\} \text{ and } \{\eta_{2t}\} \text{ are dependent such that } E[h_{2m}^{(0)}(x_1, \eta_2^{(m)})] \neq 0,$$
$$\text{for some } x_1 \in \mathcal{R}^{d_1} \times \mathcal{R}^{d_2}.$$

Under $H_1^{(m)}$, $h_{2m}^{(0)}$ is not a degenerate kernel of order 1. Hence, the V-statistic $S_{1n}^{(0)}(m)$ cannot have the convergence rate n^{-1}, as suggested in Lemma 2(i), leading to the consistency of $S_{1n}(m)$ in detecting $H_1^{(m)}$. Similarly, we can show the consistency of $J_{1n}(M)$ in detecting the following fixed alternative:

$$H_1^{(M)} : H_1^{(m)} \text{ holds for some } m \in \{0, 1, \ldots, M\}.$$

Theorem 2. *Suppose Assumptions* 1–5 *hold. Then,*

$$(i)\quad \lim_{n\to\infty} P(n[S_{1n}(m)] > c_{m\alpha}) = 1 \ \textit{under } H_1^{(m)};$$
$$(ii)\quad \lim_{n\to\infty} P(n[J_{1n}(M)] > c_\alpha) = 1 \ \textit{under } H_1^{(M)}.$$

Note that similar results to those of Theorems 1–2 hold for $S_{2n}(m)$ and

$J_{2n}(M)$, which can be implemented in a similar way to $S_{1n}(m)$ and $J_{1n}(M)$, respectively.

4. Residual Bootstrap Approximations

In this section, we introduce a residual bootstrap method to approximate the limiting null distributions in Theorem 1. The residual bootstrap method is popular in the time series literature; see, for example, Berkowitz and Kilian (2000); Paparoditis and Politis (2003); Politis (2003), and many others. The residual bootstrap procedure we use to approximate the critical values $c_{m\alpha}$ and c_α is as follows:

Step 1 Estimate the original model (2.7), and obtain the residuals $\{\widehat{\eta}_{st}\}_{t=1}^n$.

Step 2 Generate bootstrap innovations $\{\widehat{\eta}^*_{st}\}_{t=1}^n$ (after standardization) by resampling with replacement from the empirical residuals $\{\widehat{\eta}_{st}\}_{t=1}^n$.

Step 3 Given $\widehat{\theta}_{sn}$ and $\{\widehat{\eta}^*_{st}\}_{t=1}^n$, generate the bootstrap data set $\{Y^*_{st}\}_{t=1}^n$, according to

$$Y^*_{st} = f_s(\widehat{I}^*_{st-1}, \widehat{\theta}_{sn}, \widehat{\eta}^*_{st}),$$

where $\widehat{I}^*_{st}$ is the bootstrap observable information set up to time t, conditional on some assumed initial values.

Step 4 Based on $\{Y^*_{st}\}_{t=1}^n$, compute $\widehat{\theta}^*_{sn}$ in the same way as $\widehat{\theta}_{sn}$, and then calculate the corresponding bootstrap residuals $\{\widehat{\eta}^{**}_{st}\}_{t=1}^n$, with $\widehat{\eta}^{**}_{st} := g_s(Y^*_{st}, \widehat{I}^*_{st-1}, \widehat{\theta}^*_{sn})$.

Step 5 Calculate the bootstrap test statistic $S^{**}_{1n}(m)$ and $J^{**}_{1n}(M)$ in the same way as (2.11) and (2.13), respectively, where $\widehat{\eta}^{**}_{st}$ replaces $\widehat{\eta}_{st}$.

Step 6 Repeat steps 1–5 B times to obtain $\{n[S^{**}_{1nb}(m)]; b = 1, 2, \ldots, B\}$ and $\{n[J^{**}_{1nb}(M)]; b = 1, 2, \ldots, B\}$. Then, choose their αth upper quantiles, denoted by $c^*_{m\alpha}$ and c^*_α, as the approximations of $c_{m\alpha}$ and c_α, respectively.

In order to prove the validity of the bootstrap procedure in steps 1–6, we need some further notation. Let

$$h^{(0*)}_{2m}(x_1, x_2) = E^*\left[h^{(0)}_m(x_1, x_2, \widehat{\eta}^{(m*)}_3, \widehat{\eta}^{(m*)}_4)\right], \tag{4.1}$$

$$\Lambda^{(23*)}_m = E^*\left[h^{(23)}_m(\widehat{\varsigma}^{(m*)}_1, \widehat{\varsigma}^{(m*)}_2, \widehat{\varsigma}^{(m*)}_3, \widehat{\varsigma}^{(m*)}_4)\right], \tag{4.2}$$

where $\widehat{\eta}_t^{(m*)} = (\widehat{\eta}_{1t}^*, \widehat{\eta}_{2t+m}^*)$ and $\varsigma_t^{(m*)} = (\widehat{\eta}_{1t}^*, \partial g_{1t}(\widehat{\theta}_{1n})/\partial\theta_1, \widehat{\eta}_{2t+m}^*, \partial g_{2t+m}(\widehat{\theta}_{2n}) /\partial\theta_2)$. Furthermore, let $\zeta_{sn}^* = \widehat{\theta}_{sn}^* - \widehat{\theta}_{sn}$, and $\varpi_n := \{Y_{11}, Y_{12}, \ldots, Y_{1n}, Y_{21}, Y_{22}, \ldots, Y_{2n}\}$ be the given sample. Denote by E^* the expectation conditional on ϖ_n, and let $o_p^*(1)(O_p^*(1))$ be a sequence of random variables converging to zero (bounded) in probability, conditional on ϖ_n.

Because $\{\widehat{\eta}_{st}^*\}_{t=1}^N$ is an i.i.d sequence conditional on ϖ_n, a similar argument to that in Lemma 1 implies

$$S_{1n}^{**}(m) = S_{1n}^{(0*)}(m) + \zeta_{1n}^{*T} S_{1n}^{(11*)}(m) + \zeta_{2n}^{*T} S_{1n}^{(12*)}(m) + \frac{1}{2}\zeta_{1n}^{*T} S_{1n}^{(21*)}(m)\zeta_{1n}^* + \frac{1}{2}\zeta_{2n}^{*T} S_{1n}^{(22*)}(m)\zeta_{2n}^* + \zeta_{1n}^{*T} S_{1n}^{(23*)}(m)\zeta_{2n}^* + R_{1n}^*(m), \tag{4.3}$$

where $S_{1n}^{(0*)}(m)$, $S_{1n}^{(ab*)}(m)$, and $R_{1n}^*(m)$ are defined in the same way as $S_{1n}^{(0)}(m)$, $S_{1n}^{(ab)}(m)$, and $R_{1n}(m)$, respectively, with $\eta_t^{(m)}$ and $\varsigma_t^{(m)}$ replaced by $\widehat{\eta}_t^{(m*)}$ and $\widehat{\varsigma}_t^{(m*)}$, respectively. Moreover, by a similar argument to that in Lemma 1(i), we obtain

$$N[S_{1n}^{(0*)}(m)] = \sum_{j=1}^{\infty} \lambda_{jm}^* \left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi_{jm}^*(\widehat{\eta}_i^{(m*)})\right] + o_p^*(1), \tag{4.4}$$

where $E^*\Phi_{jm}^*(\widehat{\eta}_1^{(m*)}) = 0$ for all $j \geq 1$, and $E^*[\Phi_{jm}^*(\widehat{\eta}_1^{(m*)})\Phi_{j'm}^*(\widehat{\eta}_1^{(m*)})] = 1$ if $j = j'$, and 0 if $j \neq j'$.

Next, we give two technical assumptions.

Assumption 6. *The bootstrap estimator $\widehat{\theta}_{sn}^*$ satisfies that*

$$\sqrt{n}(\widehat{\theta}_{sn}^* - \widehat{\theta}_{sn}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \pi_s(Y_{st}^*, \widehat{I}_{st-1}, \widehat{\theta}_{sn}) + o_p^*(1) =: \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \pi_{st}^* + o_p^*(1),$$

where π_s is defined as in Assumption 3*, and $E^*(\pi_{st}^* | \widehat{I}_{st-1}^*) = 0$.*

Assumption 7. *The following convergence results hold:*

$$(i) \quad \frac{1}{n} \sum_{i=1}^{n} E^*[\pi_{si}^* \pi_{s'i}^{*T}] \to_p E\left[\pi_{s1}\pi_{s'1}^T\right];$$

$$(ii) \quad \frac{1}{N} \sum_{i=1}^{N} E^*[\Phi_{jm}^*(\widehat{\eta}_i^{(m*)})\pi_{si}^*] \to_p E[\Phi_{jm}(\eta_1^{(m)})\pi_{s1}],$$

as $n \to \infty$, *for* $s, s' = 1, 2$, $j \geq 1$, *and* $m = 0, 1, \ldots, M$.

Assumptions 6 and 7 are standard in proving the validity of bootstrap procedures, and they are similar to those in Assumption A7 of Escanciano (2006). For the (quasi) MLE, LSE, and NLSE or, more generally, estimators resulting from a martingale estimating equation (see Heyde (1997)), the function $\pi_s(\cdot)$ required in Assumption 6 can be expressed as $\pi_s(Y_{st}, I_{st-1}, \theta_s) = \varrho_1(\eta_{st}(\theta_s)) \times \varrho_2(I_{st-1}, \theta_s)$, for some functions $\varrho_1(\cdot)$ and $\varrho_2(\cdot)$ with $E(\varrho_1(\eta_{st}(\theta_{s0}))) = 0$. Then, in those cases, Assumptions 6 and 7 are satisfied under some mild conditions on the function $\varrho_2(\cdot)$. Note that the calculation of the bootstrap estimator $\widehat{\theta}^*_{sn}$ in step 4 may be time-consuming for some times series models (e.g., multivariate ARCH-type models) when n is large. In view of Assumption 6, we suggest generating $\widehat{\theta}^*_{sn}$ as follows:

$$\widehat{\theta}^*_{sn} = \widehat{\theta}_{sn} + \frac{1}{n} \sum_t \pi_s(Y^*_{st}, \widehat{I}^*_{st-1}, \widehat{\theta}_{sn}).$$

This saves a significant amount of computation time. In Section 5, we will apply this method to conditional variance models, and find that it generates precise critical values $c_{m\alpha}$ and c_α for the proposed HSIC-based tests.

The following theorem gives the asymptotics of our bootstrapped test statistics.

Theorem 3. *Suppose Assumptions* 1–5 *and* 6–7 *hold. Then, conditional on* ϖ_n, *(i)* $n[S^{**}_{1n}(m)] = O^*_p(1)$ *for* $0 \leq m \leq M$; *(ii)* $n[J^{**}_{1n}(M)] = O^*_p(1)$; *moreover, under* H_0,

$$(iii) \quad n[S^{**}_{1n}(m)] \to_d \chi_m \quad \text{for } 0 \leq m \leq M,$$
$$(iv) \quad n[J^{**}_{1n}(M)] \to_d \sum_{m=0}^{M} \chi_m$$

in probability as $n \to \infty$, *where* χ_m *is defined as in Theorem* 1.

By Theorem 3(i), we know that conditional on ϖ_n, our bootstrapped critical values $c^*_{m\alpha}$ and c^*_α are always bounded in probability. Under the alternative hypothesis, the proof of Theorem 2 shows that $n[S_{1n}(m)]$ and $n[J_{1n}(M)]$ converge to infinity. Therefore, the events $\{n[S_{1n}(m)] > c^*_{m\alpha}\}$ and $\{n[J_{1n}(M)] > c^*_\alpha\}$ happen with probability one for large n. This implies that our bootstrapped critical values $c^*_{m\alpha}$ and c^*_α are valid under the alternative hypothesis, although the explicit distributions of the bootstrapped test statistics are absent, and might be derived under some higher-order conditions in future.

As shown in Theorem 3(ii), the explicit distributions of the bootstrapped test statistics are the same as those of the related limiting null distributions. Hence, our bootstrapped critical values $c^*_{m\alpha}$ and c^*_{α} are also valid under the null hypothesis.

5. Simulation Studies

In this section, we compare the performance of our HSIC-based tests $S_{sn}(m)$ and $J_{sn}(M)$ ($s = 1, 2$ hereafter) with some well-known existing tests in finite samples. Below, we compute $S_{sn}(m)$ and $J_{sn}(M)$, where k and l are the Gaussian kernels and $\sigma = 1$. Additional simulation results can be found in the Supplementary Material, where k and l are chosen as inverse multi-quadratics kernels.

5.1. Conditional mean models

We generate 1,000 replications of sample size n from the following conditional mean models:

$$\begin{cases} Y_{1t} = \begin{pmatrix} \theta_{1,10} & \theta_{1,20} \\ \theta_{1,30} & \theta_{1,40} \end{pmatrix} Y_{1t-1} + \eta_{1t}, \\ Y_{2t} = \begin{pmatrix} \theta_{2,10} & \theta_{2,20} \\ \theta_{2,30} & \theta_{2,40} \end{pmatrix} Y_{2t-1} + \eta_{2t}, \end{cases} \tag{5.1}$$

where $\theta_{i0} = (\theta_{i,10}, \theta_{i,20}, \theta_{i,30}, \theta_{i,40})$ (for $i = 1, 2$) contains all unknown parameters, and $\{\eta_{1t}\}$ and $\{\eta_{2t}\}$ are sequences of i.i.d. random vectors. To generate $\{\eta_{1t}\}$ and $\{\eta_{2t}\}$, we need an auxiliary sequence of i.i.d. multivariate normal random vectors $\{u_t\}$ with mean zero, where $u_t = (u_{1t}, u_{2t}, u'_{3t}, u'_{4t})'$, with $u_{1t}, u_{2t} \in \mathcal{R}$ and $u_{3t}, u_{4t} \in \mathcal{R}^{2\times 1}$, and covariance matrix given by

$$\Omega = \begin{pmatrix} \Omega_1 & 0_{2\times 2} & 0_{2\times 2} \\ 0_{2\times 2} & \Omega_2 & \Omega_4 \\ 0_{2\times 2} & \Omega'_4 & \Omega_3 \end{pmatrix},$$

with

$$\Omega_\tau = \begin{pmatrix} 1 & \rho_\tau \\ \rho_\tau & 1 \end{pmatrix} \text{ for } \tau = 1, 2, 3, \text{ and } \Omega_4 = \begin{pmatrix} \rho_4 & \rho_4 \\ \rho_4 & \rho_4 \end{pmatrix}.$$

Here, we take $\theta_{10} = (0.4, 0.1, -1, 0.5)$, $\theta_{20} = (-1.5, 1.2, -0.9, 0.5)$, $\rho_2 = 0.5$, and $\rho_3 = 0.75$, as in El Himdi and Roy (1997), who considered the same models as those in (5.1).

Based on $\{u_t\}$, we consider six error-generating processes (EGPs):

$$\text{EGP 1}: \eta_{1t} = u_{3t}, \eta_{2t} = u_{4t} \text{ and } \rho_4 = 0;$$
$$\text{EGP 2}: \eta_{1t} = u_{3t}, \eta_{2t} = u_{4t} \text{ and } \rho_4 = 0.3;$$
$$\text{EGP 3}: \eta_{1t} = \frac{u_{1t}^2 + 1}{\sqrt{6}} u_{3t}, \eta_{2t} = |u_{1t}| u_{4t} \text{ and } \rho_4 = 0;$$
$$\text{EGP 4}: \eta_{1t} = \frac{u_{1t}^2 + 1}{\sqrt{6}} u_{3t}, \eta_{2t} = |u_{1t+3}| u_{4t} \text{ and } \rho_4 = 0;$$
$$\text{EGP 5}: \eta_{1t} = \frac{u_{1t}^2 + 1}{\sqrt{6}} u_{3t}, \eta_{2t} = |u_{2t}| u_{4t}, \rho_1 = 0.8 \text{ and } \rho_4 = 0;$$
$$\text{EGP 6}: \eta_{1t} = u_{1t} u_{3t}, \eta_{2t} = u_{2t} u_{4t}, \rho_1 = 0.8 \text{ and } \rho_4 = 0.$$

Clearly, each entry of η_{1t} or η_{2t} has mean zero and variance one. Let $\rho_{\eta_1,\eta_2}(d)$ be the cross-correlation matrix between η_{1t} and η_{2t+d}. EGP 1 is designed for the null hypothesis, because $\rho_{\eta_1,\eta_2}(d) = 0_{2\times 2}$ for all d in this case. EGPs 2–6 are set for the alternative hypotheses, because they pose a linear or non-linear dependence structure between η_{1t} and η_{2t}. Specifically, a linear dependence structure between η_{1t} and η_{2t} exists in EGP 2, with $\rho_{\eta_1,\eta_2}(d) = 0.3I_2$ for $d = 0$, and 0 otherwise. A non-linear dependence structure between η_{1t} and η_{2t} is induced by the co-factor u_{1t} in EGP 3, the lagged co-factors u_{1t} and u_{1t+3} in EGP 4, and two correlated co-factors u_{1t} and u_{2t} in EGPs 5–6. In EGPs 3–6, η_{1t} and η_{2t} are dependent, but uncorrelated.

For each replication, we fit two models in (5.1) using the LSE method. Denote by $\{\widehat{\eta}_{1t}\}$ and $\{\widehat{\eta}_{2t}\}$ the residuals of the respective fitted models. Based on $\{\widehat{\eta}_{1t}\}$ and $\{\widehat{\eta}_{2t}\}$, we compute $S_{sn}(m)$ and $J_{sn}(M)$ (S_{sn} and J_{sn} in short). The critical values of all HSIC-based tests are obtained using the residual bootstrap method with $B = 1{,}000$ in Section 4.

We also compute the test statistics $G_{sn}(M)$ (G_{sn} in short) in El Himdi and Roy (1997), and the test statistics $W_{sn}(h)$ (W_{sn} in short) in Bouhaddioui and Roy (2006), where

$$G_{1n}(M) = \sum_{m=-M}^{M} \widehat{Z}_n(m), \quad G_{2n}(M) = \sum_{m=-M}^{M} \left[\frac{n}{n - |m|}\right] \widehat{Z}_n(m),$$
$$W_{1n}(h) = \frac{\sum_{m=1-n}^{n-1} [\overline{K}(m/h)]^2 \widetilde{Z}_n(m) - d_1 d_2 A_{1n}(h)}{\sqrt{2 d_1 d_2 B_{1n}(h)}},$$
$$W_{2n}(h) = \frac{\sum_{m=1-n}^{n-1} [\overline{K}(m/h)]^2 \widetilde{Z}_n(m) - h d_1 d_2 A_1}{\sqrt{2 h d_1 d_2 B_1}}.$$

Here, $\widehat{Z}_n(m) = n[vec(R_{12}(m))]^T[R_{22}^{-1}(0) \otimes R_{11}^{-1}(0)][vec(R_{12}(m))]$,

$$R_{ij}(m) = D[(\widehat{r}_{ii}(0))^{-1/2}]\widehat{r}_{ij}(m)D[(\widehat{r}_{jj}(0))^{-1/2}],$$

$\widehat{r}_{ij}(m)$ is the sample cross-covariance matrix between $\{\widehat{\eta}_{it}\}$ and $\{\widehat{\eta}_{jt+m}\}$, $\widetilde{Z}_n(m)$ is defined in the same way as $\widehat{Z}_n(m)$, with $\widehat{\eta}_{st}$ replaced by $\widetilde{\eta}_{st}$, $\widetilde{\eta}_{st}$ is the residual from a fitted VAR(p) model for Y_{st}, $\overline{K}(\cdot)$ is a kernel function, h denotes the bandwidth, $A_1 = \int_{-\infty}^{\infty}[\overline{K}(z)]^2 dz$, $B_1 = \int_{-\infty}^{\infty}[\overline{K}(z)]^4 dz$, and

$$A_{1n}(h) = \sum_{m=1-n}^{n-1}\left(1 - \frac{|m|}{n}\right)\left[\overline{K}\left(\frac{m}{h}\right)\right]^2,$$
$$B_{1n}(h) = \sum_{m=1-n}^{n-1}\left(1 - \frac{|m|}{n}\right)\left(1 - \frac{|m|+1}{n}\right)\left[\overline{K}\left(\frac{m}{h}\right)\right]^4.$$

Note that G_{1n} is to test the cross-correlation between η_{1t} and η_{2t}; G_{2n} is its modified version for small n; W_{1n} has the same goal as G_{1n}, but with the ability to detect the cross-correlation beyond lag M; W_{2n} is the modified version of W_{1n}. Under certain conditions, the limiting null distribution of G_{1n} or G_{2n} is $\chi^2_{(2M+1)d_1d_2}$, and that of W_{1n} or W_{2n} is $N(0,1)$.

In all simulation studies, we set $m = 0$ and 3 for the single HSIC-based tests $S_{sn}(m)$, and set $M = 3$ and 6 for the joint HSIC-based test $J_{sn}(M)$. Because $S_{1n}(0) = S_{2n}(0)$, the results of $S_{2n}(0)$ are absent. For $G_{sn}(M)$, we choose $M = 3, 6$, and 9. For $W_{sn}(h)$, we follow Hong (1996) to choose $p = 3$ (or 6) when $n = 100$ (or 200), and use the kernel function $\overline{K}(z) = \sin(\pi z)/(\pi z)$ (Daniel kernel) with bandwidth $h = h_1, h_2$, or h_3, where $h_1 = [\log(n)], h_2 = [3n^{0.2}]$, and $h_3 = [3n^{0.3}]$. The significance level α is set to $1\%, 5\%$, or 10%.

Table 1 reports the power of the tests based on the two models in (5.1). The sizes of all tests correspond to those in EGP 1. From this table, our findings are as follows.

(i) The sizes of all single HSIC-based tests S_{sn} are close to their nominal values in most cases, whereas the sizes of other tests are a little unsatisfactory. For instance, J_{sn} are slightly oversized, especially at $\alpha = 5\%$ and 10%, and W_{1n} (or W_{2n}) is slightly oversized (or undersized) when $n = 200$ (or 100), at all levels. The size performance of G_{sn} depends on M: a larger value of M leads to a more undersized behavior, especially at $\alpha = 10\%$, although, in general, G_{2n} performs better than G_{1n}.

(ii) In all examined cases, the single HSIC-based test $S_{1n}(0)$ is the most powerful of the tests in EGPs 2–3 and 5–6, and the single HSIC-based test $S_{2n}(3)$ has a significant power advantage in EGP 4. These results are expected, because $S_{1n}(0)$ and $S_{2n}(3)$ are designed to examine the dependence specifically at lags 0 and 3, respectively, reflecting the setup of each EGP. Note that our HSIC-based tests in EGP 3 are more powerful than those in EGP 5. This is consistent with our setting that the dependence between η_{1t} and η_{2t} in EGP 3 is stronger than that in EGP 5.

(iii) For the linear dependence case (i.e., EGP 2), the joint HSIC-based tests J_{sn} have a comparable power performance as G_{sn}. In addition, they are much less powerful than $W_{1n}(h_1)$, but are much more powerful than $W_{2n}(h_3)$ when $n = 100$. For the non-linear dependence case (i.e., EGPs 3–6), the joint HSIC-based tests J_{sn} are, in general, much more powerful than the tests G_{sn} and W_{sn}, especially when $n = 200$. The only exception is J_{1n} in EGP 4, which cannot detect the dependence between η_{1t+m} and η_{2t} at lag $m = 3$. In contrast, J_{2n} performs very well here.

(iv) In all examined cases, the power of J_{sn} and G_{sn} decreases as the value of M increases; this tendency is vague for W_{sn}.

Overall, our single HSIC-based tests are powerful in detecting dependence at specific lags, and our joint HSIC-based tests exhibit a significant power advantage in detecting non-linear dependence, which cannot be examined easily using other tests.

5.2. Conditional variance models

We generate 1,000 replications of sample size n from the following conditional variance models:

$$\begin{cases} Y_{1t} = V_{1t}^{1/2}\eta_{1t} \quad \text{and} \quad V_{1t} = (v_{1t,ij})_{i,j=1,2}, \\ Y_{2t} = V_{2t}^{1/2}\eta_{2t} \quad \text{and} \quad V_{2t} = (v_{2t,ij})_{i,j=1,2}, \\ \text{with} \\ \begin{pmatrix} v_{1t,11} \\ v_{1t,22} \\ v_{1t,12} \end{pmatrix} = \begin{pmatrix} \theta_{1,10} + \theta_{1,20}v_{1t-1,11} + \theta_{1,30}Y_{1t-1,1}^2 \\ \theta_{1,40} + \theta_{1,50}v_{1t-1,22} + \theta_{1,60}Y_{1t-1,2}^2 \\ \theta_{1,70}\sqrt{v_{1t-1,11}v_{1t-1,22}} \end{pmatrix}, \\ \begin{pmatrix} v_{2t,11} \\ v_{2t,22} \\ v_{2t,12} \end{pmatrix} = \begin{pmatrix} \theta_{2,10} + \theta_{2,20}v_{2t-1,11} + \theta_{2,30}Y_{2t-1,1}^2 \\ \theta_{2,40} + \theta_{2,50}v_{2t-1,22} + \theta_{2,60}Y_{2t-1,2}^2 \\ \theta_{2,70}\sqrt{v_{2t-1,11}v_{2t-1,22}} \end{pmatrix}, \end{cases} \tag{5.2}$$

Table 1. Empirical sizes and power (× 100) of all tests based on the models in (5.1)

	EGP 1						EGP 2						EGP 3					
	$n = 100$			$n = 200$			$n = 100$			$n = 200$			$n = 100$			$n = 200$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$S_{1n}(0)$	0.7	5.1	11.7	1.6	5.2	11.7	47.1	69.1	79.9	85.5	95.2	97.4	80.2	94.5	97.9	99.3	100	100
$S_{1n}(3)$	0.6	5.4	11.4	0.7	4.3	10.9	1.1	5.5	13.0	0.6	4.9	9.9	0.8	5.1	10.6	1.1	5.9	10.0
$S_{2n}(3)$	1.2	5.6	12.1	1.3	4.6	9.9	1.0	5.1	11.4	1.5	5.3	9.9	1.0	5.5	11.2	0.8	4.1	9.1
$J_{1n}(3)$	0.7	5.3	12.3	1.2	5.2	11.5	19.4	44.5	58.4	55.1	78.4	85.4	30.7	64.4	79.9	88.0	96.8	98.8
$J_{1n}(6)$	0.9	6.2	14.6	1.1	6.1	13.6	12.5	32.4	48.2	40.3	66.1	76.8	11.6	37.0	55.7	66.4	89.0	95.1
$J_{2n}(3)$	1.4	7.1	12.5	1.8	6.7	13.9	19.3	42.2	57.4	54.8	78.3	87.0	31.9	61.7	77.6	86.7	96.8	98.3
$J_{2n}(6)$	1.1	6.8	13.2	1.7	6.5	12.1	13.2	32.9	47.3	38.3	62.7	76.6	10.4	36.9	56.0	66.0	87.5	94.1
$G_{1n}(3)$	0.5	3.6	7.6	0.7	5.0	10.1	17.3	41.5	57.1	69.1	88.4	93.0	10.9	23.9	33.4	14.7	29.3	39.4
$G_{1n}(6)$	0.4	2.8	7.8	0.6	4.2	9.6	17.3	41.5	57.1	43.5	70.9	83.5	5.3	14.6	24.9	8.5	21.6	32.8
$G_{1n}(9)$	0.4	1.5	4.9	0.2	3.3	6.8	8.1	25.0	39.1	29.4	55.1	69.3	2.9	10.0	16.6	6.3	17.0	25.2
$G_{2n}(3)$	0.9	4.2	8.6	0.7	5.5	10.5	18.3	43.3	59.4	69.5	89.0	93.6	11.9	25.2	35.5	15.2	29.9	40.7
$G_{2n}(6)$	0.6	4.6	10.4	1.0	5.4	10.9	12.5	30.3	45.0	45.8	72.8	84.4	6.6	18.4	29.6	10.2	24.4	34.8
$G_{2n}(9)$	0.7	4.1	9.1	0.6	4.5	9.5	7.9	25.4	36.6	34.1	60.2	74.7	5.0	15.7	23.8	8.3	19.9	28.8
$W_{1n}(h_1)$	0.9	5.2	9.4	2.2	6.9	12.8	45.6	64.9	75.2	87.5	93.9	96.9	24.2	37.4	46.9	27.2	42.4	51.1
$W_{1n}(h_2)$	0.8	4.3	8.4	1.7	6.3	12.4	30.3	53.0	65.7	78.3	89.4	93.4	18.8	30.3	39.4	21.4	36.9	46.0
$W_{1n}(h_3)$	1.0	5.4	9.4	1.6	5.4	12.5	19.6	44.5	57.3	59.6	80.2	88.0	12.6	25.3	35.5	15.1	29.4	39.6
$W_{2n}(h_1)$	0.6	4.2	7.6	2.1	6.2	11.7	41.1	62.4	72.9	86.1	93.2	96.5	21.6	35.6	44.3	25.7	40.9	50.0
$W_{2n}(h_2)$	0.4	3.2	5.6	1.4	5.0	9.8	23.1	46.4	59.4	74.3	87.7	92.1	14.7	26.2	34.3	19.2	33.5	43.5
$W_{2n}(h_3)$	0.3	1.7	4.9	0.9	3.3	6.8	11.0	28.5	43.3	49.5	73.8	83.0	8.2	17.9	24.9	10.3	22.8	31.7

	EGP 4						EGP 5						EGP 6					
	$n = 100$			$n = 200$			$n = 100$			$n = 200$			$n = 100$			$n = 200$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$S_{1n}(0)$	0.4	4.4	10.1	0.6	4.1	9.5	23.7	50.5	65.2	58.7	84.0	91.9	36.8	64.3	76.3	77.2	91.9	95.7
$S_{1n}(3)$	0.4	3.7	7.9	0.4	3.9	9.5	0.5	4.2	9.2	0.7	4.3	9.4	0.5	3.1	7.8	0.8	4.7	9.8
$S_{2n}(3)$	75.5	92.0	96.3	99.2	99.9	100	0.7	1.0	3.5	3.0	4.5	9.1	0.4	3.0	7.6	0.6	4.4	8.4
$J_{1n}(3)$	0.3	2.6	6.5	0.4	2.7	7.8	4.5	23.6	34.4	20.7	46.3	60.4	7.6	25.3	41.9	35.8	63.8	75.7
$J_{1n}(6)$	0.3	1.7	5.2	0.2	2.1	5.3	1.3	9.5	19.3	9.0	28.8	45.4	1.7	12.4	25.5	17.9	40.5	57.5
$J_{2n}(3)$	28.4	57.2	76.2	86.7	96.5	98.5	4.7	21.5	32.4	19.3	45.7	59.7	5.6	23.6	38.8	35.4	63.0	75.9
$J_{2n}(6)$	9.7	34.3	53.7	64.4	88.1	94.6	1.9	8.5	19.4	8.8	27.5	45.9	1.8	10.3	23.4	11.3	22.9	31.9
$G_{1n}(3)$	10.4	21.4	31.9	12.8	27.1	38.4	5.5	14.7	23.7	8.1	19.6	28.0	3.9	12.7	20.3	4.9	14.2	24.8
$G_{1n}(6)$	4.6	13.7	21.4	8.4	19.8	30.2	2.0	9.6	16.7	3.9	14.2	24.6	2.8	8.8	15.2	2.9	10.6	16.3
$G_{1n}(9)$	2.9	8.3	15.4	5.4	15.6	24.5	1.4	5.3	12.3	2.7	10.6	17.5	1.7	6.9	11.2	2.1	7.9	13.9
$G_{2n}(3)$	12.3	24.7	35.5	13.8	28.6	39.7	6.1	15.9	25.3	8.3	20.2	29.4	4.2	13.7	22.9	5.0	14.6	25.5
$G_{2n}(6)$	7.0	17.8	26.8	9.0	22.9	32.6	3.2	12.8	21.3	4.6	16.5	26.1	3.7	11.6	19.3	3.3	11.6	19.0
$G_{2n}(9)$	4.8	14.6	25.8	7.0	19.6	27.9	2.6	11.1	19.5	4.5	13.0	22.5	3.1	10.4	18.7	2.7	9.8	17.6
$W_{1n}(h_1)$	2.8	9.6	16.5	6.6	15.7	24.8	14.1	20.5	34.1	16.0	28.3	35.7	11.6	21.7	30.8	11.3	22.9	31.9
$W_{1n}(h_2)$	7.9	16.9	25.1	10.9	23.6	34.1	10.5	19.2	29.4	12.9	23.5	34.2	8.1	17.4	27.0	8.8	18.3	27.6
$W_{1n}(h_3)$	8.7	18.2	27.1	10.7	25.9	35.7	6.9	18.2	26.2	9.2	19.9	29.6	6.7	15.9	24.1	5.5	15.1	21.8
$W_{2n}(h_1)$	2.3	8.2	14.1	6.3	14.8	23.4	13.2	19.9	32.1	15.5	26.9	34.2	10.0	19.7	22.6	10.5	21.9	30.2
$W_{2n}(h_2)$	6.3	13.6	20.1	9.2	20.6	30.4	8.2	16.5	23.6	11.7	20.7	31.6	6.5	13.9	20.6	7.2	16.5	24.0
$W_{2n}(h_3)$	5.6	11.8	17.5	8.3	18.2	29.1	4.0	10.8	17.5	6.5	15.1	21.3	3.2	9.3	15.4	3.6	10.3	16.9

† For W_{sn}, $h_1 = [\log(n)]$, $h_2 = [3n^{0.2}]$ and $h_3 = [3n^{0.3}]$

where $\theta_{i0} = (\theta_{i,10}, \theta_{i,20}, \ldots, \theta_{i,70})$ (for $i = 1, 2$) contains all unknown parameters, and $\{\eta_{1t}\}$ and $\{\eta_{2t}\}$ are sequences of i.i.d. random vectors generated as in (5.1). In (5.2), two CC-MGARCH models are studied, as in Tse (2002). Following Tse (2002), we set $\theta_{10} = (0.2, 0.5, 0.1, 0.2, 0.5, 0.1, 0.5)$ and $\theta_{20} = (0.3, 0.4, 0.2, 0.3, 0.4, 0.2, 0.6)$. For each replication, we fit the models in (5.2) using the Gaussian-QMLE method. Denote by $\{\widehat{\eta}_{1t}\}$ and $\{\widehat{\eta}_{2t}\}$ the residuals from the respective fitted models. Based on $\{\widehat{\eta}_{1t}\}$ and $\{\widehat{\eta}_{2t}\}$, we compute $S_{sn}(m)$ and $J_{sn}(M)$, and their critical values as before.

We also compute the test statistics $L_{sn}(M)$ and $T_{sn}(M)$ (L_{sn} and T_{sn}, in short) of Tchahou and Duchesne (2013), where

$$L_{1n}(M) = \sum_{m=-M}^{M} n\rho^2_{\widehat{q}_{1t},\widehat{q}_{2t}}(m), \quad L_{2n}(M) = \sum_{m=-M}^{M} \left[\frac{n^2}{n-|m|}\right] \rho^2_{\widehat{q}_{1t},\widehat{q}_{2t}}(m),$$

$$T_{1n}(M) = \sum_{m=-M}^{M} n \cdot tr(C_{12}^T(m)C_{11}^{-1}(0)C_{12}(m)C_{22}^{-1}(0)),$$

$$T_{2n}(M) = \sum_{m=-M}^{M} \left[\frac{n^2}{n-|m|}\right] \cdot tr(C_{12}^T(m)C_{11}^{-1}(0)C_{12}(m)C_{22}^{-1}(0)).$$

Here, $\rho_{\widehat{q}_{1t},\widehat{q}_{2t}}(m)$ is the sample cross-correlation between $\{\widehat{q}_{1t}\}$ and $\{\widehat{q}_{2t+m}\}$, $C_{ij}(m)$ is the sample cross-covariance matrix between $\{\widehat{\varphi}_{it}\}$ and $\{\widehat{\varphi}_{jt+m}\}$, $\widehat{q}_{st} = \widehat{\eta}_{st}^T\widehat{\eta}_{st}$, and $\widehat{\varphi}_{st} = vech(\widehat{\eta}_{st}\widehat{\eta}_{st}^T)$. Note that L_{1n} (or T_{1n}) is used to test the cross-correlation between two transformed (or original) residuals, and L_{2n} (or T_{2n}) is its modified version for small n. Under certain conditions, the limiting null distribution of L_{1n} or L_{2n} is $\chi^2_{(2M+1)}$, and that of T_{1n} or T_{2n} is $\chi^2_{(2M+1)d_1^*d_2^*}$, where $d_s^* = d_s(d_s+1)/2$ for $s = 1, 2$.

In all simulation studies, we choose the values of m and M as in the previous subsection. The significance level α is set to 1%, 5%, or 10%. Table 2 summarizes the power results based on the two models in (5.2). The sizes of all tests correspond to those in EGP 1. From this table, our findings are as follows.

(i) The sizes of all tests are close to their nominal values, although most T_{sn} are slightly oversized.

(ii) Similarly to the results shown in Table 1, the single HSIC-based test $S_{1n}(0)$ or $S_{1n}(3)$, as expected, is the most powerful of the tests, and the HSIC-based tests in EGP 3 are more powerful than those in EGP 5.

(iii) For the linear dependence case (i.e., EGP 2), all joint HSIC-based tests J_{sn}

are more powerful than L_{sn} and T_{sn}. For the non-linear dependence case (i.e., EGP 3–6), all J_{sn} have larger power than L_{sn} and T_{sn} in most cases, but this advantage is small, especially for $J_{sn}(6)$. There are two exceptions in which some J_{sn} exhibit low power: first, $J_{1n}(3)$ and $J_{1n}(6)$ have no power in EGP 4 (see also Table 1); second, $J_{2n}(6)$ is less powerful than most L_{sn} and T_{sn}, especially for $n = 200$. Because the cross-correlation between η_{1t}^2 and η_{2t}^2 is high in EGPs 2–6, the relatively good power performance of L_{sn} and T_{sn} in some cases is not unexpected.

(iv) The power of the tests J_{sn}, L_{sn}, and T_{sn} decreases as the value of M increases in all examined cases.

Overall, our single HSIC-based tests exhibit good power in detecting dependence at specific lags, and our joint HSIC-based tests could be more powerful than other tests in detecting either linear or non-linear dependence. Moreover, our additional simulation results in the Supplementary Material indicate that the selection of the kernel function could affect the performance of our HSIC-based tests, although the overall patterns of performance are similar. Hence, choosing kernel functions optimally based on some criteria is important in practice and deserves future investigation.

6. A Real Example

In this section, we study two bivariate time series. The first consists of index series from the Russian market and the Indian market: the Russia Trading System Index (RTSI) and the Bombay Stock Exchange Sensitive Index (BSESI), respectively. The second includes two Chinese indices: the ShangHai Securities Composite index (SHSCI) and the ShenZhen Index (SZI). The data were measured each day (Monday to Friday), from October 8, 2014 to September 29, 2017. The final sample comprised 1,088 days. Missing data due to holidays were removed before the analysis, after which the final data set includes $n = 672$ daily observations. The resulting four time series are denoted by $\{\text{RTSI}_t; t = 1, \ldots, n\}$, $\{\text{BSESI}_t; t = 1, \ldots, n\}$, $\{\text{SHSCI}_t; t = 1, \ldots, n\}$, and $\{\text{SZI}_t; t = 1, \ldots, n\}$, respectively.

As usual, we consider the log-return of each data set:

$$Y_{1t} = \begin{pmatrix} Y_{1t,1} \\ Y_{1t,2} \end{pmatrix} = \begin{pmatrix} \log(\text{RTSI}_t) - \log(\text{RTSI}_{t-1}) \\ \log(\text{BSESI}_t) - \log(\text{BSESI}_{t-1}) \end{pmatrix},$$

Table 2. Empirical sizes and power (×100) of all tests based on the models in (5.2)

	EGP 1						EGP 2						EGP 3					
	$n=200$			$n=300$			$n=200$			$n=300$			$n=200$			$n=300$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$S_{1n}(0)$	0.7	4.3	10.5	1.6	5.4	9.2	100	100	100	100	100	100	100	100	100	100	100	100
$S_{1n}(3)$	1.2	5.2	11.0	0.5	5.1	10.1	1.3	5.8	10.8	1.5	5.8	9.6	0.8	4.1	8.9	0.8	5.4	10.8
$S_{2n}(3)$	1.1	4.5	9.3	0.6	4.6	9.7	0.9	5.1	9.3	0.9	4.6	9.3	1.2	4.9	9.5	1.2	4.5	8.6
$J_{1n}(3)$	0.7	4.5	10.7	0.8	4.7	9.0	99.2	99.9	99.9	100	100	100	97.7	99.6	99.8	100	100	100
$J_{1n}(6)$	0.7	3.7	9.1	0.4	4.1	8.8	91.3	98.5	99.4	99.8	100	100	85.9	96.5	98.6	99.2	100	100
$J_{2n}(3)$	0.8	4.1	9.2	1.0	5.5	11.6	98.6	99.8	99.9	100	100	100	97.8	99.6	100	100	100	100
$J_{2n}(6)$	0.6	4.0	9.0	1.0	4.9	10.3	91.0	97.8	99.1	99.9	100	100	83.8	96.4	98.8	95.5	95.9	96.0
$L_{1n}(3)$	1.2	3.9	9.9	1.3	6.1	10.0	15.7	34.8	46.3	32.2	54.3	65.4	87.6	91.2	92.7	92.4	94.4	95.0
$L_{1n}(6)$	1.1	4.3	9.2	0.9	5.6	11.3	8.5	25.2	37.7	22.0	41.5	54.8	82.0	88.4	90.7	90.0	92.4	93.2
$L_{1n}(9)$	0.9	3.6	9.2	1.1	4.5	9.5	9.5	18.8	30.8	15.8	35.3	47.9	78.2	85.2	88.2	88.4	91.5	92.3
$L_{2n}(3)$	1.2	4.1	10.1	1.3	6.2	10.3	16.0	35.2	46.6	32.4	54.5	65.5	87.6	91.2	92.7	92.4	94.4	95.0
$L_{2n}(6)$	1.5	5.2	10.5	1.0	5.8	12.1	9.0	26.0	38.7	22.6	42.0	55.5	82.4	88.5	90.8	90.0	92.4	93.2
$L_{2n}(9)$	0.9	4.4	11.5	1.3	4.8	10.5	6.1	20.5	32.3	16.9	36.7	49.2	78.6	85.8	88.6	88.4	91.6	92.4
$T_{1n}(3)$	2.1	6.7	11.9	2.2	6.4	11.6	39.5	60.4	70.1	61.7	77.4	84.5	79.5	85.6	87.4	87.0	90.4	92.1
$T_{1n}(6)$	1.7	6.5	11.6	1.6	6.2	11.4	26.3	41.5	54.3	45.9	63.1	72.7	68.3	76.5	79.3	77.9	83.5	86.5
$T_{1n}(9)$	1.3	5.8	10.8	1.2	4.8	9.9	14.8	31.2	41.6	32.3	53.7	64.4	60.7	70.7	74.9	72.2	78.4	81.4
$T_{2n}(3)$	2.2	7.4	12.8	2.3	6.7	12.7	41.0	60.8	70.9	61.5	78.0	84.5	79.9	85.7	87.8	87.2	91.0	92.1
$T_{2n}(6)$	2.2	7.8	13.4	2.0	7.5	12.5	25.1	45.9	57.7	47.5	64.5	74.3	69.3	77.4	80.3	78.6	83.9	87.2
$T_{2n}(9)$	2.6	7.5	13.5	1.5	7.0	12.5	18.4	36.7	48.3	35.3	58.0	68.0	63.6	73.2	76.4	73.8	79.4	82.1

	EGP 4						EGP 5						EGP 6					
	$n=200$			$n=300$			$n=200$			$n=300$			$n=200$			$n=300$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$S_{1n}(0)$	0.5	3.7	7.7	0.5	4.4	9.7	76.3	89.4	94.4	92.1	98.5	99.3	92.4	97.8	99.1	98.8	99.8	99.8
$S_{1n}(3)$	1.0	4.3	8.9	1.0	4.1	10.1	0.6	3.9	9.0	0.7	4.9	9.1	0.8	4.5	10.3	1.0	4.5	10.1
$S_{2n}(3)$	100	100	100	100	100	100	0.7	4.7	9.2	0.6	5.2	9.2	0.7	3.5	7.8	0.6	4.6	9.5
$J_{1n}(3)$	0.3	2.5	6.5	0.7	3.9	8.6	33.9	61.2	73.5	61.8	82.0	88.9	56.4	80.2	88.0	86.3	95.3	97.9
$J_{1n}(6)$	0.3	1.3	4.1	0.3	3.4	7.0	13.6	40.2	56.6	38.1	64.0	76.6	30.5	57.8	72.2	66.8	85.3	93.0
$J_{2n}(3)$	97.1	99.4	99.8	100	100	100	30.1	61.3	74.7	62.0	81.2	89.0	56.6	78.8	87.5	85.9	95.1	98.1
$J_{2n}(6)$	83.1	97.0	98.4	99.8	100	100	12.8	38.2	55.3	36.7	63.5	77.1	27.8	57.8	71.7	64.7	84.6	91.9
$L_{1n}(3)$	86.6	91.2	92.1	93.2	94.4	95.1	51.9	61.1	70.2	66.7	76.4	80.9	49.6	64.8	73.4	68.1	79.5	85.3
$L_{1n}(6)$	80.7	87.2	89.4	90.7	93.2	94.3	42.7	57.3	64.3	57.3	69.5	75.6	41.0	57.1	64.1	58.4	72.9	79.0
$L_{1n}(9)$	75.1	84.1	86.1	87.9	91.8	92.8	37.6	52.2	59.1	51.6	63.8	70.0	31.8	51.8	59.1	52.7	67.8	74.9
$L_{2n}(3)$	87.0	91.4	92.3	93.2	94.4	95.1	52.0	61.2	71.3	66.7	76.5	81.5	49.7	65.0	73.5	68.1	79.6	85.5
$L_{2n}(6)$	81.3	87.4	89.7	90.7	93.2	94.3	43.3	58.3	65.0	57.6	69.7	75.8	41.6	57.1	64.5	58.5	73.0	79.1
$L_{2n}(9)$	76.6	84.8	87.4	88.0	91.9	93.0	38.1	52.9	60.3	52.0	64.1	70.7	33.1	53.1	60.5	53.4	68.5	75.5
$T_{1n}(3)$	80.5	85.6	88.1	88.1	90.5	92.2	51.7	59.8	64.4	58.1	67.5	72.2	43.8	55.1	61.2	56.2	65.5	70.1
$T_{1n}(6)$	67.2	75.6	79.3	79.8	85.4	87.8	43.2	52.3	57.1	48.2	60.1	65.3	34.7	45.8	52.7	44.5	55.7	61.8
$T_{1n}(9)$	60.4	69.0	72.6	71.7	78.5	82.1	37.7	46.7	52.1	41.7	51.8	57.3	29.3	40.4	46.2	40.1	50.4	55.8
$T_{2n}(3)$	86.6	91.2	92.1	88.1	90.7	92.3	52.0	59.2	65.2	58.9	67.7	72.6	44.4	55.1	62.8	56.7	65.7	70.4
$T_{2n}(6)$	68.7	77.2	81.2	81.0	86.3	88.2	44.9	53.3	57.8	49.5	60.9	66.4	36.9	47.4	54.3	45.7	57.3	62.5
$T_{2n}(9)$	63.6	70.8	76.0	73.3	79.9	82.9	40.1	49.0	55.3	43.5	53.8	58.9	32.2	43.7	49.6	42.0	52.5	59.0

Table 3. Estimation results for both fitted BEKK models

Parameters	Estimates		Parameters	Estimates	
A_1	$\hat{a}_{1,11}$	0.2832×10^{-3}	A_2	$\hat{a}_{2,11}$	0.2528×10^{-5}
	$\hat{a}_{1,12}$	0.0050×10^{-3}		$\hat{a}_{2,12}$	0.3856×10^{-5}
	$\hat{a}_{1,22}$	0.0022×10^{-3}		$\hat{a}_{2,22}$	0.6714×10^{-5}
B_{11}	$\hat{b}_{11,11}$	0.4662	B_{21}	$\hat{b}_{21,11}$	0.3098
	$\hat{b}_{11,22}$	-0.0619		$\hat{b}_{21,22}$	0.3195
B_{12}	$\hat{b}_{12,11}$	-0.1149	B_{22}	$\hat{b}_{22,11}$	-0.1264
	$\hat{b}_{12,22}$	0.3357		$\hat{b}_{22,22}$	-0.0692
C_{11}	$\hat{c}_{11,11}$	0.3569	C_{21}	$\hat{c}_{21,11}$	0.6808
	$\hat{c}_{11,22}$	0.2222		$\hat{c}_{21,22}$	0.6783
C_{12}	$\hat{c}_{12,11}$	0.5370	C_{22}	$\hat{c}_{22,11}$	0.6431
	$\hat{c}_{12,22}$	0.9027		$\hat{c}_{22,22}$	0.6455

† Note that A_s is a symmetric matrix, and all B_{sj} and C_{sj} are diagonal matrixes.

$$Y_{2t} = \begin{pmatrix} Y_{2t,1} \\ Y_{2t,2} \end{pmatrix} = \begin{pmatrix} \log(\mathrm{SHSCI}_t) - \log(\mathrm{SHSCI}_{t-1}) \\ \log(\mathrm{SZI}_t) - \log(\mathrm{SZI}_{t-1}) \end{pmatrix}.$$

An analysis of the ACF and PACF of $Y_{1t,1}, Y_{1t,2}, Y_{2t,1}, Y_{2t,2}$, and their squares indicates they have no conditional mean structure, but they do have a conditional variance structure. Motivated by this, we use the following BEKK model with the Gaussian-QMLE method to fit Y_{1t} and Y_{2t}:

$$\begin{aligned} Y_{st} &= \Sigma_{st}^{1/2}\eta_{st}, \\ \Sigma_{st} &= A_s + B_{s1}^T Y_{1t-1} Y_{1t-1}^T B_{s1} + \cdots + B_{sp}^T Y_{1t-p} Y_{1t-p}^T B_{sp} \\ &\quad + C_{s1}^T \Sigma_{st-1} C_{s1} + \cdots + C_{sq}^T \Sigma_{st-q} C_{sq}, \end{aligned}$$

for $s = 1, 2$, where $A_s = C_{s0}^T C_{s0}$, with C_{s0} being a triangular 2×2 matrix, and $B_{s1}, \ldots, B_{sp}, C_{s1}, \ldots, C_{sq}$ are 2×2 diagonal matrices. Table 3 reports the estimates for both fitted models. The respective p-values of portmanteau tests $Q(3)$, $Q(6)$, and $Q(9)$ of Ling and Li (1997) are $0.7698, 0.5179$, and 0.5967 for Y_{1t}, and $0.5048, 0.7328$, and 0.8746 for Y_{2t}. This implies that both fitted BEKK models are adequate.

Next, we apply our joint HSIC-based tests $J_{sn}(M)$ to check whether Y_{1t} and Y_{2t} behave independently of each other. As a comparison, we also consider the tests $L_{sn}(M)$ and $T_{sn}(M)$. Table 4 reports the p-values for all six tests. Here, except for $J_{2n}(M)$, with $M \geq 7$, all examined joint HSIC-based tests $J_{sn}(M)$ convey strong evidence that Y_{1t} and Y_{2t} are not independent. However, neither $L_{sn}(M)$ nor $T_{sn}(M)$ achieves this for $M \geq 2$.

Table 4. The p-value for all six joint tests up to lag $M = 0, 1, \ldots, 10$.

M	Tests					
	J_{1n}	J_{2n}	L_{1n}	L_{2n}	T_{1n}	T_{2n}
0	0.0000	0.0000	0.0134	0.0134	0.0000	0.0000
1	0.0000	0.0000	0.0428	0.0428	0.0125	0.0124
2	0.0000	0.0000	**0.0881**	**0.0879**	**0.1965**	**0.1956**
3	0.0000	0.0260	**0.0610**	**0.0605**	**0.1055**	**0.1035**
4	0.0000	0.0040	**0.1137**	**0.1128**	**0.2979**	**0.2927**
5	0.0090	0.0240	**0.2111**	**0.2095**	**0.4640**	**0.4557**
6	0.0230	0.0280	**0.2762**	**0.2739**	**0.5958**	**0.5851**
7	0.0220	**0.0720**	**0.3315**	**0.3282**	**0.7093**	**0.6972**
8	0.0280	**0.0730**	**0.4079**	**0.4037**	**0.6708**	**0.6540**
9	0.0450	**0.0830**	**0.4491**	**0.4437**	**0.7645**	**0.7475**
10	0.0230	**0.1040**	**0.5761**	**0.5706**	**0.8359**	**0.8199**

† A p-value larger than 5% is in boldface.

To get more information, we further plot the values of the single version of J_{sn}, L_{1n}, and T_{1n} in Fig 1. That is, Fig 1 plots the values of $S_{sn}(m)$, $L_{1n,s}(m)$, and $T_{1n,s}(m)$, for $m \geq 0$, where

$$L_{1n,1}(m) = n\rho^2_{\widehat{q}_{1t},\widehat{q}_{2t}}(m), \;\; L_{1n,2}(m) = n\rho^2_{\widehat{q}_{1t},\widehat{q}_{2t}}(-m),$$
$$T_{1n,1}(m) = n \cdot tr(C_{12}^T(m)C_{11}^{-1}(0)C_{12}(m)C_{22}^{-1}(0)),$$
$$T_{1n,2}(m) = n \cdot tr(C_{12}^T(-m)C_{11}^{-1}(0)C_{12}(-m)C_{22}^{-1}(0)),$$

and all notation is inherited from Section 5.2. The limiting null distribution of $L_{1n,s}(m)$ is χ_1^2, and that of $T_{1n,s}(m)$ is χ_9^2. Similarly to $S_{sn}(m)$, $L_{1n,s}(m)$ and $T_{1n,s}(m)$ capture the linear dependence between η_{1t} and η_{1t+m} at specific lag m. The corresponding single version results for L_{2n} and T_{2n} are similar to those for L_{1n} and T_{1n}; hence, they are not displayed here.

From Fig 1, we first find that all single tests indicate a strong contemporaneously causal relationship between the Chinese market and the Russian and Indian (R&I) market. Second, $S_{1n}(1)$ implies that the R&I market has a significant affect on the Chinese market one day later. However, according to $S_{2n}(3)$ (or $S_{2n}(10)$), the impact of the Chinese market on the R&I market appears after three (or ten) days. These findings demonstrate an asymmetric causal relationship between two markets. Because none of the examined $L_{1n,s}(m)$ and $T_{1n,s}(m)$ can detect a causal relationship for $m \geq 1$, the contemporaneous causal relationship mainly causes the significance of $L_{sn}(1)$ and $T_{sn}(1)$ in Table 4, and the lagged causal relationship may be non-linear. Because the R&I market has a

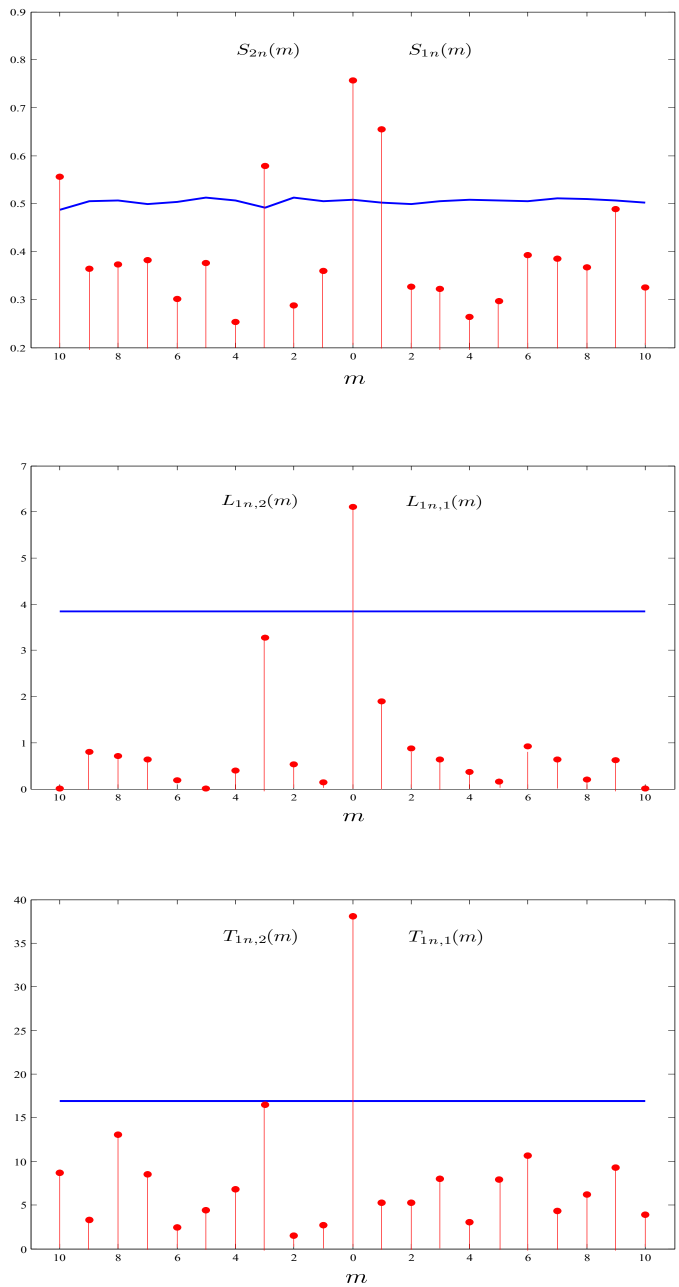

Figure 1. The values of single tests $S_{1n}(m)$, $L_{1n,1}(m)$, and $T_{1n,1}(m)$ (right panel) across m, and the values of single tests $S_{2n}(m)$, $L_{1n,2}(m)$, and $T_{1n,2}(m)$ (left panel) across m. The solid lines are 95% one-sided confidence bounds of the tests.

higher degree of globalization and marketization, it may have a faster impact on other economies. In contrast, the Chinese market is more localized, and its influence on other economies tends to be slower, but can last much longer. This long-term effect may be caused by "the Belt and Road Initiatives" of the Chinese government, implemented in 2015. Hence, the asymmetric phenomenon between two markets seems reasonable, and may help the government to formulate efficient policy, and investors to design more useful investment strategies.

7. Conclusion

We have applied the HSIC principle to derive novel one-sided omnibus tests for detecting independence between two multivariate stationary time series. The resulting HSIC-based tests have a non-degenerate asymptotical representation under the null hypothesis, and are shown to be consistent. A residual bootstrap method is used to obtain the critical values for our HSIC-based tests, and its validity is justified. Unlike existing cross-correlation-based tests for linear dependence, our HSIC-based tests look for general dependence between two unobservable innovation vectors. Hence, they can provide researchers with information that is more complete on the causal relationship between two time series. The importance of our HSIC-based tests is illustrated by simulation results and a real-data analysis. The generality of the HSIC method means that our methodology may be applied to many other important testing problems, such as testing for model adequacy (Davis et al. (2018)), testing for independence among multi-dynamic systems (Pfister et al. (2018)), and testing for independence in high-dimensional systems (Yao, Zhang and Shao (2018)). We leave these interesting topics to future study.

Supplementary Material

The online Supplementary Material contains additional simulation results and the proofs of all lemmas and theorems.

Acknowledgments

The authors greatly appreciate the helpful comments and suggestions of the anonymous referees, Associate Editor, and Co-Editor. Wang's work is supported by Guangdong Basic and Applied Basic Research Foundation (2020A1515010 821) and Fundamental Research Funds for the Center University (12619624). Li's work is supported by RGC of Hong Kong (No. 17304417). Zhu's work is supported by RGC of Hong Kong (Nos. 17306818 and 17305619), NSFC (Nos. 11571348, 11690014, 11731015 and 71532013), Seed Fund for Basic Research (No. 201811159049), Hung Hing Ying Physical Sciences Research Fund 2017-18, and Fundamental Research Funds for the Central University (19JNYH08).

References

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**, 79–109.

Berkowitz, J. and Kilian, L. (2000). Recent developments in bootstrapping time series. *Econometric Reviews* **19**, 1–48.

Bouhaddioui, C. and Roy, R. (2006). A generalized portmanteau test for independence of two infinite-order vector autoregressive series. *Journal of Time Series Analysis* **27**, 505–544.

Cheung, Y.-W. and Ng, L. K. (1996). A causality-in-variance test and its application to financial market prices. *Journal of Econometrics* **72**, 33–48.

Choudhry, T., Papadimitriou, F. I. and Shabi, S. (2016). Stock market volatility and business cycle: Evidence from linear and nonlinear causality tests. *Journal of Banking & Finance* **66**, 89–101.

Comte, F. and Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* **84**, 61–84.

Davis, R.A., Matsui, M., Mikosch, T. and Wan, P. (2018) Applications of distance correlation to time series. *Bernoulli* **24**, 3087–3116.

Denker, M. and Keller, G. (1983) On U-statistics and v. Mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **64**, 505–522.

Diks, C. and Wolski, M. (2016). Nonlinear granger causality: Guidelines for multivariate analysis. *Journal of Applied Econometrics* **31**, 1333–1351.

Dunford, N. and Schwartz, J. T. (1963). *Linear Operators Part 2: Spectral Theory.* Interscience, New York.

El Himdi, K. and Roy, R. (1997). Tests for noncorrelation of two multivariate ARMA time series. *Canadian Journal of Statistics* **25**, 233–256.

Engle, R. F. and Kroner, F. K. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11** 122–150.

Escanciano, J. C. (2006). Goodness-of-fit tests for linear and non-linear time series models. *Journal of the American Statistical Association* **101**, 531–541.

Fokianos, K. and Pitsillou, M. (2017). Consistent testing for pairwise dependence in time series. *Technometrics* **59**, 262–270.

Francq, C. and Zakoïan, J. M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, Chichester, UK.

Gretton, A., Bousquet, O., Smola, A.J. and Scholkopf, B. (2005). Measuring statistical dependence with hilbert-schmidt norms. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT)*, 63–77.

Gretton, A., Fukumizu, K., Teo, C.H., Song, L., Schoumlkopf, B. and Smola, A. (2008). A kernel statistical test of independence. *Advances in the 20th International Conference on Neural Information Procesing Systems* 20, MIT Press, 585–592.

Gretton, A. and Györfi, L. (2010). Consistent nonparametric tests of independence. *Journal of Machine Learning Research* **11**, 1391–1423.

Hafner, C. M. and Preminger, A. (2009). On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* **100**, 2044–2054.

Hallin, M. and Saidi, A. (2005). Testing non-correlation and non-causality between multivariate ARMA time series. *Journal of Time Series Analysis* **26**, 83–106.

Hallin, M. and Saidi, A. (2007). Optimal tests of non-correlation between multivariate time series. *Journal of the American Statistical Association* **102**, 938–952.

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. *Journal of the American Statistical Association* **71**, 378–385.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Applications*. Springer-Verlag, Berlin.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance* **49**, 1639–1664.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19**, 293–325.

Hong, Y. (1996). Testing for independence between two covariance stationary time series. *Biometrika* **83**, 615–625.

Hong, Y. (2001a). A test for volatility spillover with application to exchange rates. *Journal of Econometrics* **103**, 183–224.

Hong, Y. (2001b). Testing for independence between two stationary time series via the empirical characteristic function. *Annals of Economics and Finance* **2**, 123–164.

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

Lee, T.-H. and Long, X. (2009). Copula-based multivariate GARCH model with uncorrelated dependent errors. *Journal of Econometrics* **150**, 207–218.

Ling, S. and Li, W. K. (1997). Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis* **18**, 447–464.

Ling, S. and McAleer, M. (2003). Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* **19**, 280–310.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.

Paparoditis, E. and Politis, D. N. (2003) Residual-based block bootstrap for unit root testing. *Econometrica* **71**, 813–855.

Peters, J. (2008). *Asymmetries of Time Series under Inverting Their Direction*. Diploma Thesis, University of Heidelberg.

Pfister, N., Bühlmann, P., Scholkopf, B. and Peters, J. (2018). Kernel-based tests for joint independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 5–31.

Pham, D., Roy, R. and Cédras, L. (2003). Tests for non-correlation of two cointegrated ARMA time series. *Journal of Time Series Analysis* **24**, 553–577.

Pierce, A. (1977). Lack of dependence among economic variables. *Journal of the American Statistical Association* **72**, 11–22.

Politis, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science* **18**, 219–230.

Robbins, M. W. and Fisher, T. J. (2015). Cross-correlation matrices for tests of independence and causality between two multivariate time series. *Journal of Business & Economic Statistics* **33**, 459–473.

Schwert, G. W. (1979). Tests of causality: the message in the innovations. 55–96 in Karl Brunner and Allan H. Meltzer (eds.), Three Aspects of Policy and Policymaking: Knowledge, Data, and Institutions. Amsterdam: North-Holland.

Sejdinovic, D., Sriperumbudur, A., Gretton, A. and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics* **41**, 2263–2291.

Sen, A. and Sen, B. (2014). On testing independence and goodness-of-fit in linear models. *Biometrika* **101**, 927–942.

Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. In: Handbook of Financial Time Series (T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch, eds.) 201–229. Springer, Berlin.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica* **48**, 1–48.

Shao, X. (2009). A generalized portmanteau test for independence between two stationary time series. *Econometric Theory* **25**, 195–210.

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35**, 2769–2794.

Tchahou, H. N. and Duchesne, P. (2013). On testing for causality in variance between two multivariate time series. *Journal of Statistical Computation and Simulation* **83**, 2064–2092.

Tsay, R. S. (2014). *Multivariate Time Series Analysis: with R and Financial Applications.* John Wiley&Sons, New York.

Tse, Y. K. (2002). Residual-based diagnostics for conditional heteroscedasticity models. *Econometrics Journal* **5**, 358–374.

Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business & Economic Statistics* **20**, 351–362.

Wang, Y., Wu, C. and Yang, L. (2013). Oil price shocks and stock market activities: Evidence from oil-importing and oil-exporting countries. *Journal of Comparative Economics* **41**, 1220–1239.

Yao, S., Zhang, X. and Shao, X. (2018). Testing mutual independence in high dimension via distance covariance. *Journal of Royal Statistical Society: Series B (Statistical Methodology)* **80**, 455–480.

Zhang, Q., Filippi, S., Gretton, A. and Sejdinovic, D. (2018). Large-scale kernel methods for independence testing. *Statistics and Computing* **28**, 113-130.

Zhang, X., Song, L., Gretton, A. and Smola, A. J. (2009). Kernel measures of independence for non-iid data. In *Advances in Neural Information Processing Systems*, 1937–1944.

Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis* **33**, 438–457.

Guochang Wang

College of Economics, Jinan University, Guangzhou, China.

E-mail: twanggc@jnu.edu.cn

Wai Keung Li

Department of Mathematics and Information Technology, The Education University of Hong

Kong, Hong Kong.
E-mail: waikeungli@eduhk.hk
Ke Zhu
Department of Statistics & Actuarial Science, The University of Hong Kong, Hong Kong.
E-mail: mazhuke@hku.hk

(Received April 2018; accepted March 2019)