Econometric Theory, **34**, 2018, 815–849.
doi:10.1017/S026646661700010X

CHARACTERISTIC FUNCTION BASED TESTING FOR CONDITIONAL INDEPENDENCE: A NONPARAMETRIC REGRESSION APPROACH

XIA WANG
Sun Yat-Sen University

YONGMIAO HONG
Cornell University
Xiamen University

We propose a characteristic function based test for conditional independence, applicable to both cross-sectional and time series data. We also derive a class of derivative tests, which deliver model-free tests for such important hypotheses as omitted variables, Granger causality in various moments and conditional uncorrelatedness. The proposed tests have a convenient asymptotic null $N(0,1)$ distribution, and are asymptotically locally more powerful than a variety of related smoothed nonparametric tests in the literature. Unlike other smoothed nonparametric tests for conditional independence, we allow nonparametric estimators for both conditional joint and marginal characteristic functions to jointly determine the asymptotic distributions of the test statistics. Monte Carlo studies demonstrate excellent power of the tests against various alternatives. In an application to testing Granger causality, we document the existence of nonlinear relationships between money and output, which are missed by some existing tests.

Wang acknowledges financial supports from the National Science Foundation of China (No. 71401160), the Ministry of Education of Humanities and Social Sciences Project of China (No. 14YJC790120), Fujian Provincial Key Laboratory of Statistics (Xiamen University, No. 2016003). We thank the editor Peter C. B. Phillips, the co-editor Liangjun Su, two referees, and seminar participants at Chinese Academy of Sciences, Kansas Econometrics Workshop, New York Econometrics Camp, Singapore Management University, and UC Riverside for their comments and suggestions. We are most grateful to Peter C. B. Phillips, who has gone beyond the call of duty for an editor in carefully correcting our English. We also thank Liangjun Su for providing us the computer codes to implement his tests. All remaining errors are solely our own. Address correspondence to Yongmiao Hong, Department of Economics and Department of Statistical Sciences, Cornell University, 424 Uris Hall, Ithaca, NY 14850, USA, and Wang Yanan Institute for Studies in Economics, Economic Building, Xiamen University, Xiamen 361005, China; e-mail: yh20@cornell.edu; and Xia Wang, Lingnan (University) College, Sun Yat-Sen University, Guangzhou 510275, China.

1. INTRODUCTION

Conditional independence is a widely maintained condition and encompasses many important hypotheses in econometrics and statistics (Dawid, 1979, 1980). Let X, Y, and Z be random vectors. Then Y is said to be independent of Z given X, denoted as $Y \perp Z|X$, if the joint distribution function of (Y, Z) conditional on X is equal to the product of the conditional marginal distribution functions of Y and Z.

To motivate, we provide a few examples. The first example is the Markov property. A time series $\{X_t\}$ is said to be a Markov process of order one if $X_{t+1} \perp (X_{t-1}, X_{t-2}, \ldots)|X_t$. For a Markov process, the current variable X_t contains all useful information in predicting the future behavior of $\{X_t\}$. This property is broadly used in economics and finance (e.g., Easley and O'Hara, 1987; Rust, 1994). When it holds, one can capture the full dynamics of $\{X_t\}$ by using a time series model with one lag only.

The second example is non-Granger causality (Granger, 1969, 1980). For two time series $\{Z_t\}$ and $\{Y_t\}$, $\{Z_t\}$ does not Granger-cause $\{Y_t\}$ in distribution if $Y_t \perp Z_{t-q}^{t-1}|Y_{t-p}^{t-1}$, where $Z_{t-q}^{t-1} = (Z_{t-1}, \ldots, Z_{t-q})$, $Y_{t-p}^{t-1} = (Y_{t-1}, \ldots, Y_{t-p})$, and p, q are lag orders. If this hypothesis is rejected, Z_{t-q}^{t-1} is useful in predicting the future distribution of $\{Y_t\}$. Granger (1969) proposes an F test for Granger causality in a linear regression setup, which is widely applied in empirical studies. However, this test may miss important nonlinear phenomena, such as the asymmetric effect of monetary policies (Kim and Nelson, 2006) and the asymmetric behavior of asset returns (Campbell, 1992; Peiró, 1999).

The third example is the missing at random property, which is often assumed in treatment response analysis (e.g., Hahn, Todd, and Klaauw, 2001; Wang, Linton, and Hardle, 2004). A sample is said to be missing at random if missingness does not depend on the conditioning variables in the sample (Rubin, 1976). Specifically, suppose Y is an outcome, X is a covariate, and Z is a binary indicator for treatment, which is equal to 1 if Y is observed and 0 otherwise. Then Y is missing at random conditional on X, if $Y \perp Z|X$. When this assumption holds, one can obtain consistent estimation by throwing away the unobservable subsample and can point-identify the treatment effect in response analysis. However, abuse of this assumption may render inconsistent estimation, which is called selectivity bias in the literature (Heckman, 1976; Little, 1985). Horowitz and Manski (2000) and Manski (2000, 2003, 2007) show that without this assumption, one can only obtain interval estimation for the treatment effect rather than point-identify it.

The last example is exogeneity. Suppose $Y = g(X, U)$, where $g(\cdot, \cdot)$ is an unknown function, X is an observed covariate, and U is an unobservable error. Then X is exogenous if $X \perp U$. To test exogeneity, researchers (e.g., Blundell and Horowitz, 2007; Lee, 2013) introduce an instrumental variable Z for X and show that X is exogenous if and only if $Y \perp Z|X$. Exogeneity is fundamental to econometric modeling and inference. Models that suffer from endogeneity problems require different estimation methods, which are usually less efficient than those when X is exogenous.

Motivated by widespread applications, a growing literature focuses on testing the hypothesis of conditional independence. Su and White (2007, 2008) develop nonparametric tests based on some weighted distances between characteristic functions and between densities respectively. Song (2009) proposes a test for conditional independence between two continuous random variables based on the Rosenblatt transforms. Huang (2010) develops a nonparametric test using a maximal nonlinear conditional correlation. Su and White (2012) test conditional independence via a local polynomial quantile regression. Bouezmarni, Rombouts, and Taamouti (2012) and Taamouti, Bouezmarni, and El Ghouch (2014) develop nonparametric copula-based tests for conditional independence and non-Granger causality in distribution respectively. Bouezmarni and Taamouti (2012) test conditional independence by comparing conditional distribution functions, and Linton and Gozalo (2014) propose a test using the empirical distribution function. Finally, Su and White (2014) propose two smoothed empirical likelihood ratio tests, and Huang, Sun, and White (2016) develop an integrated conditional moment test based on a distance between restricted and unrestricted probability measures.

In this paper, we propose a new characteristic function-based test for conditional independence using a nonparametric regression approach. The proposed test has the following features.

First, the test can detect a class of local alternatives that converge to the null hypothesis at a faster rate than existing smoothed nonparametric tests for conditional independence in the literature. Let d_x, n and $h = h(n)$ denote the dimension of X, the sample size and the bandwidth, and we test whether $Y \perp Z | X$ holds. Then the rate of local alternatives for the test is $n^{-1/2}h^{-d_x/4}$, which is faster than the rate of local alternatives for such nonparametric tests as Su and White (2007, 2008, 2014), Huang (2010), and Bouezmarni et al. (2012). The latter all depend on the dimensions of other variables as well as d_x.

Second, by differentiating the proposed omnibus test statistic, we obtain a class of derivative tests that can be used to gauge patterns of conditional dependence, including model-free tests for omitted variables, Granger causality in various moments, and conditional uncorrelatedness. The derivative test for omitted variables is asymptotically more powerful than the smoothed nonparametric tests of Fan and Li (1996), Lavergne and Vuong (2000), and Aït-Sahalia, Bickel, and Stoker (2001).

Third, unlike other smoothed nonparametric tests for conditional independence, we use a single bandwidth in estimating both conditional joint and marginal characteristic functions. The corresponding nonparametric estimators jointly determine the asymptotic distribution of the test statistic. This avoids the delicate business of choosing multi-bandwidths, and results in a significantly better size for the test in finite samples due to fewer negligible higher order terms.

Finally, the test applies to both cross-sectional and time series data, and has a convenient asymptotic null $N(0, 1)$ distribution. We require the conditioning

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

vector X to have a continuous distribution, but allow Y and Z to have either discrete or continuous distributions or a mixture of them.

In Section 2, we state the hypothesis of conditional independence. In Section 3, we construct the omnibus test statistic. We then derive the asymptotic distribution of the test statistic in Section 4 and its asymptotic local power in Section 5. Section 6 develops a class of derivative tests. In Section 7, we study the finite sample performance of the test. Section 8 considers an application to testing nonlinear Granger causality between money and output. Section 9 concludes. All proofs are given in the Mathematical Appendix. Computer codes to implement the test are available from the authors upon request.

2. CONDITIONAL INDEPENDENCE AND HYPOTHESIS OF INTEREST

Let X, Y, and Z be random vectors of dimension d_x, d_y, and d_z respectively. Suppose we have an identically distributed but weakly dependent random sample $\{X_t, Y_t, Z_t\}_{t=1}^n$. Denote $f(\cdot|\cdot)$ as the conditional density (or mass) function of one random vector given another. For convenience, $f(\cdot|\cdot)$ is referred to as a conditional density below. However, we allow Y and Z to have either discrete or continuous distributions or a mixture of them. Our null hypothesis of conditional independence is

$$\mathbb{H}_0 : P\left[f(y,z|X) = f(y|X)f(z|X)\right] = 1 \text{ for any } (y,z) \in \mathbb{R}^{d_y+d_z}. \quad \textbf{(1)}$$

The alternative hypothesis is

$$\mathbb{H}_A : P\left[f(y,z|X) \neq f(y|X)f(z|X)\right] > 0 \text{ for some non-negligible values of } (y,z). \quad \textbf{(2)}$$

As the Fourier transform of $f(y,z|x)$, we can use the conditional characteristic function to represent $\mathbb{H}_0$ equivalently. Denote the conditional characteristic functions as

$$\phi_{yz}(u,v,x) = E\left[e^{\mathbf{i}(u'Y_t+v'Z_t)}|X_t = x\right],$$
$$\phi_y(u,x) = E\left(e^{\mathbf{i}u'Y_t}|X_t = x\right), \quad \phi_z(v,x) = E\left(e^{\mathbf{i}v'Z_t}|X_t = x\right),$$

where $E(\cdot|X_t = x)$ and $\text{cov}(\cdot,\cdot|X_t = x)$ denote the values of the corresponding conditional expectation and covariance evaluated at $X_t = x$. Furthermore, define a conditional generalized covariance

$$\sigma(u,v,x) = \text{cov}\left(e^{\mathbf{i}u'Y_t}, e^{\mathbf{i}v'Z_t}|X_t = x\right), \quad (u,v) \in \mathbb{R}^{d_y+d_z}. \quad \textbf{(3)}$$

Straightforward algebra shows that

$$\sigma(u,v,X_t) = \phi_{yz}(u,v,X_t) - \phi_y(u,X_t)\phi_z(v,X_t). \quad \textbf{(4)}$$

For a strictly stationary time series $\{Y_t\}$, when $Z_t = Y_{t-k}$ and $X_t = (Y_{t-1},\ldots,Y_{t-k+1})'$, $\sigma(u,v,x)$ could be viewed as a generalized partial autocovariance because its partial derivative

$$\frac{\partial^2}{\partial u \partial v}\sigma(u,v,x)|_{(u,v)=(0,0)} = -\text{cov}(Y_t, Y_{t-k}|X_t = x)$$

is the conventional partial autocovariance function (PACF). This extends the concept of generalized autocovariance function of Hong (1999). The function $\sigma(u,v,x)$ can capture any type of pairwise conditional partial dependence over various lags, including nonlinear time series with zero partial autocovariance, such as bilinear, nonlinear moving average, and ARCH/GARCH processes. To see this, we rewrite $\sigma(u,v,x)$ using the Taylor series expansion:

$$\sigma(u,v,x) = \sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\frac{(\mathbf{i}u)^m(\mathbf{i}v)^l}{m!l!}\operatorname{cov}\left(Y_t^m, Y_{t-k}^l|X_t = x\right).$$

Intuitively, when all moments of Y_t exist, testing whether $\sigma(u,v,x)=0$ is equivalent to testing whether Y_t^m and Y_{t-k}^l are partially uncorrelated for any pair of (m,l), where $m,l = 0,1,2,\ldots$.

With the definition of $\sigma(u,v,x)$, we can rewrite $\mathbb{H}_0$ and $\mathbb{H}_A$ equivalently as follows:

$$\mathbb{H}_0: \quad P[\sigma(u,v,X_t)=0]=1 \text{ for any } (u,v)\in\mathbb{R}^{d_y+d_z} \tag{5}$$

versus

$$\mathbb{H}_A: \quad P[\sigma(u,v,X_t)\neq 0]>0 \text{ for } (u,v) \text{ in some set with positive Lebesgue measure.} \tag{6}$$

It is important to emphasize that we must check (5) for all $(u,v)\in\mathbb{R}^{d_y+d_z}$ rather than only a subset of $\mathbb{R}^{d_y+d_z}$. On the other hand, by differentiating $\sigma(u,v,X_t)$ with respect to u and/or v at the origin, we can infer patterns of conditional dependence (cf. Section 6).

3. NONPARAMETRIC REGRESSION BASED TESTING

3.1. Generalized Nonparametric Regression

Given a sample $\{X_t,Y_t,Z_t\}_{t=1}^n$, we can estimate $\sigma(u,v,X_t)$ and check whether it is identically zero for all $(u,v)\in\mathbb{R}^{d_y+d_z}$. We estimate $\phi_{yz}(u,v,x)$, $\phi_y(u,x)$, and $\phi_z(v,x)$ nonparametrically. Since these functions are generalized regression functions (e.g., $\phi_{yz}(u,v,x)=E[e^{\mathbf{i}(u'Y_t+v'Z_t)}|X_t=x]$), we use a pth order local polynomial regression. Because $\phi_y(u,x)$ and $\phi_z(v,x)$ can be obtained from $\phi_{yz}(u,v,x)$ by setting $v=0$ and $u=0$ respectively, we only need to estimate $\phi_{yz}(u,v,x)$.

To estimate $\phi_{yz}(u,v,x)$, we consider the following local weighted least squares problem:

$$\min_{\beta\in\mathbb{C}^{d_x p+1}}\sum_{t=1}^{n}\left|e^{\mathbf{i}(u'Y_t+v'Z_t)}-\beta_0-\sum_{1\le|\mathbf{j}|\le p}\beta_{\mathbf{j}}'(X_t-x)^{\mathbf{j}}\right|^2 K_h(X_t-x),$$
$$\text{where } x\in\mathbb{R}^{d_x}, u\in\mathbb{R}^{d_y}, v\in\mathbb{R}^{d_z}. \tag{7}$$

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

Here, we use the notations of Masry (1996a,b) and Su and White (2012). Denote $j_1, \ldots, j_{d_x}$ as non-negative integers, $\mathbf{j} \equiv (j_1, \ldots, j_{d_x})$, $|\mathbf{j}| = \sum_{j=1}^{d_x} j_i$, $x^{\mathbf{j}} = \prod_{i=1}^{d_x} x_i^{j_i}$, $\sum_{0 \le |\mathbf{j}| \le p} \equiv \sum_{k=0}^{p} \sum_{0 \le j_1, \ldots, j_{d_x} \le k, j_1 + \cdots + j_{d_x} = k}$, $\beta_{\mathbf{j}} \equiv \beta_{\mathbf{j}}(x, u, v) \equiv \frac{1}{\mathbf{j}!} D^{|\mathbf{j}|} \phi_{yz}(u, v, x)$ with $D^{|\mathbf{j}|} \phi_{yz}(u, v, x) = \frac{\partial^{|\mathbf{j}|} \phi_{yz}(u, v, x)}{\partial^{j_1} x_1 \cdots \partial^{j_{d_x}} x_{d_x}}$ and $\mathbf{j}! = \prod_{i=1}^{d_x} j_i!$. In addition, β denotes the parameter vector formed by stacking the $\beta_{\mathbf{j}}$ vectors in lexicographic order. If we define $N_l = (l + d_x - 1)!/[l!(d_x - 1)!]$, the number of distinct d_x-tuples $\mathbf{j}$ with $|\mathbf{j}| = l$, and $N = \sum_{l=1}^{p} N_l$, then β is a $N \times 1$ vector. Moreover, $K_h(x) = h^{-d_x} K(\frac{x}{h})$, where $K : \mathbb{R}^{d_x} \to \mathbb{R}$ is a kernel and $h = h(n)$ is a bandwidth. The solution to (7) is given as

$$\hat{\beta} \equiv \hat{\beta}(u, v, x) = (X'WX)^{-1} X'WQ, \quad x \in \mathbb{R}^{d_x}, \tag{8}$$

where X is a $n \times N$ matrix formed by stacking the $(X_t - x)^{\mathbf{j}}$ vectors in lexicographic order as the tth row, $W = \text{diag}[K_h(X_1 - x), \ldots, K_h(X_n - x)]$, and $Q = \left[e^{\mathbf{i}(u'Y_1 + v'Z_1)}, \ldots, e^{\mathbf{i}(u'Y_n + v'Z_n)}\right]'$. Then $\phi(u, v, x)$ can be estimated by the local intercept estimator $\hat{\beta}_0(u, v, x)$, namely,

$$\hat{\phi}_{yz}(u, v, x) = \sum_{t=1}^{n} \hat{W}\left(\frac{X_t - x}{h}, x\right) e^{\mathbf{i}u'Y_t + \mathbf{i}v'Z_t}, \tag{9}$$

where $\hat{W}(t, x) \equiv e_1' S_n(x)^{-1} [1, t^{(p)}]' K(t)/h^{d_x}$, $e_1 = (1, 0, \ldots, 0)'$, $S_n(x) = X'WX$, and $t^{(p)}$ denotes the vector formed by stacking the $t^{\mathbf{j}}$ vectors in lexicographic order.

Arrange the N_l d_x-dimensional tuples as a sequence in a lexicographical order (with the highest priority to the last position), so that $\varphi_l(1) \equiv (0, 0, \ldots, l)$ is the first element in the sequence and $\varphi_l(N_l) \equiv (l, 0, \ldots, 0)$ is the last element. Define $\mu_{\mathbf{j}} = \int_{\mathbb{R}^{d_x}} x^{\mathbf{j}} K(x) dx$ and a $N \times N$ matrix

$$S = \begin{pmatrix} S_{0,0} & S_{0,1} & \cdots & S_{0,p} \\ S_{1,0} & S_{1,1} & \cdots & S_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p,0} & S_{p,1} & \cdots & S_{p,p} \end{pmatrix},$$

where $S_{i,j}$ is a $N_i \times N_j$ matrix whose (l, s) element is $\mu_{\varphi_i(l) + \varphi_j(s)}$. Denote $S^{-1} = (S^{i,j})_{0 \le i,j \le p}$ with $S^{i,j}$ being a $N_i \times N_j$ matrix, and X_h being a $n \times N$ matrix formed by staking the $[(X_t - x)/h]^{\mathbf{j}}$ vectors in lexicographic order as the tth row. By Fan and Gijbels (1996) and Hjellvik, Yao, and Tjøstheim (1998), we have

$$\begin{aligned} \hat{\phi}_{yz}(u, v, x) &= \frac{1}{nh^{d_x} g(x)} \sum_{t=1}^{n} \left[\sum_{l=0}^{p} S^{0,p} \left(\frac{X_t - x}{h}\right)_{(l)}\right] K\left(\frac{X_t - x}{h}\right) e^{\mathbf{i}u'Y_t + \mathbf{i}v'Z_t} [1 + o_P(1)] \\ &\equiv \frac{1}{nh^{d_x} g(x)} \sum_{t=1}^{n} D\left(p, \frac{X_t - x}{h}\right) K\left(\frac{X_t - x}{h}\right) e^{\mathbf{i}u'Y_t + \mathbf{i}v'Z_t} [1 + o_P(1)] \end{aligned} \tag{10}$$

uniformly in $x \in \mathbb{G}$, where $g(x)$ is the density of X_t and $\mathbb{G}$ is any compact subset in the interior region with $g(x) > 0$. Here, $\left(\frac{X_t - x}{h}\right)_{(l)}$ denotes the $(N_{l-1}+1)$th to the N_lth elements of the tth row of X_h. We can easily check that when $p = 1$ (the local linear case), $D(1, \frac{X_t - x}{h}) = 1$. Given (10), we obtain the conditional marginal estimators $\hat{\phi}_y(u,x) = \hat{\phi}_{yz}(u,0,x)$ and $\hat{\phi}_z(v,x) = \hat{\phi}_{yz}(0,v,x)$ immediately.

We note that the nonparametric estimators for $\phi_{yz}(u,v,x)$, $\phi_y(u,x)$, and $\phi_z(v,x)$ only involve smoothing over X_t. This differs from nonparametric estimators for $f(y,z|x)$ and $f(y|z,x)$, both of which involve smoothing over Y_t, Z_t, X_t, and their convergence rates are inversely related to $d_x + d_y + d_z$. It also differs from nonparametric estimation for $E(e^{\mathbf{i}u'Y_t}|X_t, Z_t)$, which involves smoothing over X_t and Z_t, and the convergence rate is inversely related to $d_x + d_z$. The reduction in curse of dimensionality due to the use of a regression approach makes our test asymptotically more powerful than the existing smoothed nonparametric tests of $\mathbb{H}_0$ in the literature, including those of Su and White (2007, 2008, 2014), Huang (2010) and Bouezmarni et al. (2012). See Section 5 for more discussion.

3.2. Nonparametric Based Test Statistic

Under $\mathbb{H}_0$, $\sigma(u,v,X_t) = 0$ $a.s.$ for all u, v. Therefore, we can test $\mathbb{H}_0$ by using the quadratic form

$$\hat{M} = \frac{1}{n}\sum_{t=1}^{n} \iint |\hat{\sigma}(u,v,X_t)|^2 a(X_t) dW_1(u) dW_2(v), \tag{11}$$

where $\hat{\sigma}(u,v,X_t) = \hat{\phi}_{yz}(u,v,X_t) - \hat{\phi}_y(u,X_t)\hat{\phi}_z(v,X_t)$, $a : \mathbb{G} \to \mathbb{R}^+$ is a weighting function for X_t, $W_1 : \mathbb{R}^{d_y} \to \mathbb{R}^+$ and $W_2 : \mathbb{R}^{d_z} \to \mathbb{R}^+$ are right-continuous weighting functions of u, v that weigh sets symmetric about the origin equally. The weighting function $a(\cdot)$ is commonly used in the literature (e.g., Hjellvik et al., 1998; Aït-Sahalia et al., 2001; Chen and Hong, 2010). For example, since nonparametric estimation at sparse extreme observations is inaccurate, a suitably truncated function $a(\cdot)$ can alleviate the influence of unreliable estimates, although the test with a truncated function may miss derivations from the null hypothesis in the tail distribution of X_t. The use of $W_1(u)$ and $W_2(v)$ allows us to check many points for u, v. One example is the $N(0,1)$ cumulative distribution function (CDF).

Our test statistic is a standardized version of (11), namely,

$$\widehat{SM} = \left(nh^{d_x/2}\hat{M} - \hat{B}\right)/\sqrt{\hat{V}}, \tag{12}$$

where

$$\hat{B} = h^{-d_x/2} \iint \left[1 - |\hat{\phi}_y(u,x)|^2\right] dW_1(u) \int \left[1 - |\hat{\phi}_z(v,x)|^2\right] dW_2(v) a(x) dx \times \int D^2(p,\tau) K^2(\tau) d\tau, \tag{13}$$

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

$$\hat{V} = 2\int \left[\iint |\hat{\Phi}_y(u_1+u_2,x)|^2 dW_1(u_1)dW_1(u_2) \iint |\hat{\Phi}_z(v_1+v_2,x)|^2 dW_2(v_1)dW_2(v_2) \right] a^2(x)dx \\ \times \int \left[\int D(p,\tau)D(p,\tau+\eta)K(\tau)K(\tau+\eta)d\tau \right]^2 d\eta, \quad \textbf{(14)}$$

and $\hat{\Phi}_s(u+v,x) = \hat{\phi}_s(u+v,x) - \hat{\phi}_s(u,x)\hat{\phi}_s(v,x)$ for $s = y$ or z.

The factors $\hat{B}$ and $\hat{V}$ are the estimators for the asymptotic mean and asymptotic variance of (11). When the dimensions of X, Y, or Z are high, the calculation of $\widehat{SM}$ involves high-dimensional integration. One can use numerical integration or simulation techniques.

Both $\hat{B}$ and $\hat{V}$ in (13) and (14) are derived under $\mathbb{H}_0$ as the sample size $n \to \infty$. However, they may not approximate well the mean and variance of (11) in finite samples respectively, which may lead to a poor size for the test. To fix this, we also consider a finite-sample version of $\widehat{SM}$:

$$\widehat{SM}_n = \left(nh^{d_x/2}\hat{M} - \hat{B}_n\right)/\sqrt{\hat{V}}, \quad \textbf{(15)}$$

where

$$\hat{B}_n = h^{d_x/2}\sum_{t=1}^{n}\sum_{s=1}^{n} a(X_t)\hat{W}\left(\frac{X_s - X_t}{h}, X_t\right)^2 \iint |\hat{\varepsilon}_y(u,X_s)\hat{\varepsilon}_z(v,X_s)|^2 dW_1(u)dW_2(v),$$

with $\hat{\varepsilon}_y(u,X_s) = e^{\mathbf{i}u'Y_s} - \hat{\phi}_y(u,X_s)$ and similarly for $\hat{\varepsilon}_z(v,X_s)$. Both $\hat{\varepsilon}_y(u,X_s)$ and $\hat{\varepsilon}_z(v,X_s)$ could be viewed as estimated generalized residuals. One could also replace $\hat{V}$ by a finite-sample version, but our simulations show that $\widehat{SM}_n$ in (15) has performed reasonably well in finite samples.

4. ASYMPTOTIC DISTRIBUTION

To derive the asymptotic distribution of $\widehat{SM}$ in (12) under $\mathbb{H}_0$, we first impose regularity conditions.

Assumption A.1. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. (a) The stochastic process $\{X_t, Y_t, Z_t\}$ is strictly stationary absolutely regular on $\mathbb{R}^{d_x+d_y+d_z}$ with β-mixing coefficients satisfying $\sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} < C$ for some $0 < \delta < \frac{1}{3}$; (b) the marginal density $g(x)$ of X_t is positive, bounded and continuously differentiable in $x \in \mathbb{G} \subset \mathbb{R}^{d_x}$ up to order $p+1$, where $\mathbb{G}$ is a compact support of the weight function $a(\cdot)$ defined in Assumption A.4(b) below.

Assumption A.2. Let $\phi_{yz}(u,v,x)$, $\phi_y(u,x)$, $\phi_z(v,x)$ be the conditional characteristic functions of (Y_t, Z_t), Y_t and Z_t given $X_t = x$ respectively. For each pair $(u,v) \in \mathbb{R}^{d_y+d_z}$, $\phi_{yz}(u,v,x)$, $\phi_y(u,x)$, and $\phi_z(v,x)$ are continuously differentiable with respect to $x \in \mathbb{G}$ up to order $p+1$.

Assumption A.3. $K : \mathbb{R}^{d_x} \to \mathbb{R}^+$ is a product of some univariate kernel k, i.e., $K(u) = \prod_{j=1}^{d_x} k(u_j)$, where $k : \mathbb{R} \to \mathbb{R}^+$ satisfies the Lipschitz condition and is symmetric about zero, bounded, and square-integrable with $\int_{-\infty}^{\infty} u^2 k(u)du = \mu_2 < \infty$. The functions $H_j(u) = u^j K(u)$ for all j with $0 \le |j| \le 2p+1$ are Lipschitz continuous.

Assumption A.4. (a) $W_1 : \mathbb{R}^{d_y} \to \mathbb{R}^+$ and $W_2 : \mathbb{R}^{d_z} \to \mathbb{R}^+$ are nondecreasing right-continuous functions that weigh sets symmetric about zero equally, with $\int_{\mathbb{R}^{d_y}} \|u\|^4 dW_1(u) < \infty$ and $\int_{\mathbb{R}^{d_z}} \|v\|^4 dW_2(v) < \infty$; (b) $a : \mathbb{G} \to \mathbb{R}^+$ is a bounded continuous function.

Assumptions A.1 and A.2 impose conditions on the data generating process (DGP). Assumption A.1(a) is standard for application of a central limit theorem for degenerate U statistics of a weakly dependent process (e.g., Tenreiro, 1997). The β-mixing condition restricts the degree of temporal dependence in (X_t, Y_t, Z_t), which is generally adopted in nonparametric time series analysis; see, e.g., Hjellvik et al. (1998), Su and White (2007, 2008) and Chen and Hong (2010). A variety of time series processes, such as ARMA, bilinear and ARCH, satisfy the β-mixing condition (Fan and Li, 1999). The β-mixing condition could be relaxed to the $\alpha-$mixing condition; see Su and White (2012). Assumption A.1(b) rules out discrete distributions for X_t. However, we could extend our test to cover the discrete case for X_t in a similar way to Su and White (2008). Note that the components of Y_t and Z_t can be either continuous or discrete random variables or a mixture of them, and Assumption A.2 holds if $f(y,z|x)$, $f(y|x)$, $f(z|x)$ are continuously differentiable with respect to $x \in \mathbb{G}$ up to order $p+1$.

Assumptions A.3 and A.4 impose conditions on various weighting functions. Because we use a local polynomial regression, the second order kernel is appropriate. In comparison, Su and White (2007, 2008) require using higher order kernels. Assumption A.4 imposes mild conditions on $W_1(u)$, $W_2(v)$ and $a(x)$, ensuring the existence of the integrals in (12). Many functions satisfy Assumption A.4(a), examples being the CDFs with finite fourth order moments. For convenience, we can use product forms $W_1(u) = \prod_{i=1}^{d_y} w(u_i)$ and $W_2(v) = \prod_{i=1}^{d_z} w(v_i)$, where $w(\cdot)$ is a univariate CDF.

We now derive the asymptotic distribution of $\widehat{SM}$ under $\mathbb{H}_0$.

THEOREM 1. *Suppose Assumptions A.1–A.4 hold, and the bandwidth* $h = cn^{-\lambda}$ *for* $\frac{2}{d_x+4p+4} < \lambda < \frac{2}{3d_x}$, *where* $0 < c < \infty$. *Then* $\widehat{SM} \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0$ *as* $n \to \infty$.

Theorem 1 requires $p > d_x/2 - 1$, where p is the local polynomial order. Thus, we can use the simple local linear estimation ($p = 1$) when $d_x \le 3$, but have to use a higher order local polynomial when $d_x \ge 4$. This is due to the need to control the nonparametric estimation bias. As mentioned by Li and Racine (2007, p. 87), fitting a pth order local polynomial with the second order kernel is similar to using a rth order kernel for local constant estimation, where $r = 2(\lfloor p/2 \rfloor + 1)$. We can show that our results require $r > d_x/2$ for local constant estimation. We do not impose any restriction on d_y and d_z because $\widehat{SM}$ only involves smoothing on X_t and so the convergence rate only depends on d_x. This differs from the existing nonparametric tests for conditional independence (e.g., Su and White, 2007, 2008; Huang, 2010; Bouezmarni et al., 2012), which involve smoothing of $d_x + d_y + d_z$ or at least $d_x + d_z$ dimensions. For example, Su and White's (2007) Assumption A.2 implies $d_x + d_z < 4r/3$, while Su and White (2008) require $d_x + d_y + d_z \leq 7$ even when a higher order kernel is used.

It is important to emphasize that we use the same bandwidth h in estimating $\phi_{yz}(u, v, x)$, $\phi_y(u, x)$, and $\phi_z(v, x)$. As a result, the conditional joint and marginal characteristic function estimators have the same convergence rate and jointly determine the limiting distribution of $\widehat{SM}$. In contrast, the existing smoothed tests in the literature use different bandwidths to estimate joint and marginal functions of interest (e.g., regression functions, density functions, or characteristic functions), and carefully control the relative speeds for the bandwidths so that non-parametric marginal estimators converge faster than their joint counterparts and have no impact on the asymptotic distributions of the tests. This approach is taken by Fan and Li (1996), Lavergne and Vuong (2000), Aït-Sahalia et al. (2001), Su and White (2007, 2008, 2014), and Su and Ullah (2009). However, while the marginal estimators converge faster than the joint estimators, their convergence rates may be close to each other. Thus, ignoring the impact of the marginal estimators may lead to a poor size performance in finite samples. In contrast, by using the same bandwidth, we provide a better asymptotic approximation, which yields a better size performance in finite samples, as is confirmed in our simulation study below. We also avoid the delicate business of choosing multi-bandwidths.

Theorem 1 allows a wide range of admissible rates for the bandwidth h. In practice, one could choose h via simple rules of thumb. However, it is desirable to use data-driven methods to choose h. While the bandwidth based on cross-validation is asymptotically optimal for estimation of $\sigma(u, v, x)$ in terms of the mean squared error, it may not be optimal for our test. For testing problems, the central concern is about Type I and Type II errors. In different but related contexts, Gao and Gijbels (2008) and Sun, Phillips, and Jin (2008) propose some novel methods to choose a data-driven bandwidth by considering a tradeoff between Type I and Type II errors. Specifically, based on the Edgeworth expansion of the asymptotic distribution of a test statistic under a local alternative, Gao and Gijbels (2008) choose h to maximize the power of their test subject to a control of Type I error, and Sun et al. (2008) choose h to minimize a weighted average of Type I and Type II errors. It is possible to extend these approaches to $\widehat{SM}$. Nevertheless, the analytical expressions for the leading terms of the two type errors of $\widehat{SM}$ are rather involved. This is beyond the scope of this paper and will be pursued in a subsequent study.

The $\widehat{SM}$ test applies to both cross-sectional and time series data. Under $\mathbb{H}_0$, it has an asymptotic $N(0,1)$ distribution and is asymptotically pivotal. In comparison with Su and White's (2012) test, $\widehat{SM}$ is pivotal for not only martingale difference sequence observations, but also data with weak dependence.

5. ASYMPTOTIC LOCAL POWER

To compare the relative efficiency of the $\widehat{SM}$ test with some existing nonparametric tests for conditional independence in the literature, we first consider the following class of local alternatives:

$$\mathbb{H}_1(a_n): f(y,z|x) = f(y|x)f(z|x) + a_n q(y,z|x), \tag{16}$$

where $q(y,z|x)$ is bounded and continuously differentiable up to order $p+1$ with respect to $x \in \mathbb{G}$, with $q(y,z|x) \neq 0$ and $\iint q(y,z|x)dydz = 0$ for all $x \in \mathbb{G}$. The term $a_n q(y,z|x)$ characterizes the departure from $\mathbb{H}_0$, and a_n is the rate at which the deviation vanishes to 0 as $n \to \infty$. For notational simplicity, we have suppressed the dependence of $f(y,z|x)$ on n. By Fourier transform, we obtain

$$\phi_{yz}(u,v,x) = \phi_y(u,x)\phi_z(v,x) + a_n\delta(u,v,x),$$

where $\delta(u,v,x) = \iint e^{\mathbf{i}(u'y+v'z)}q(y,z|x)dydz$, with

$$\gamma \equiv \iiint |\delta(u,v,x)|^2 dW_1(u)dW_2(v)a(x)g(x)dx < \infty.$$

THEOREM 2. *Suppose Assumptions A.1–A.4 and* $\mathbb{H}_1(a_n)$ *with* $a_n = n^{-1/2}h^{-d_x/4}$ *hold, and the bandwidth* $h = cn^{-\lambda}$ *for* $\frac{2}{d_x+4p+4} < \lambda < \frac{2}{3d_x}$, *where* $0 < c < \infty$. *Then, the power of* $\widehat{SM}$ *satisfies* $P\left[\widehat{SM} \geq z_\alpha | \mathbb{H}_1(a_n)\right] \to 1 - \Phi(z_\alpha - \gamma/\sqrt{V})$ *as* $n \to \infty$, *where* $\Phi(\cdot)$ *is the* $N(0,1)$ *CDF,* z_α *is the one sided critical value of* $N(0,1)$ *at significance level* α, *and*

$$V = 2\int \left[\iint |\Phi_y(u_1+u_2,x)|^2 dW_1(u_1)dW_1(u_2) \iint |\Phi_z(v_1+v_2,x)|^2 dW_2(v_1)dW_2(v_2)\right] a^2(x)dx \times \int \left[\int D(p,\tau)D(p,\tau+\eta)K(\tau)K(\tau+\eta)d\tau\right]^2 d\eta, \tag{17}$$

with $\Phi_s(u+v,x) = \phi_s(u+v,x) - \phi_s(u,x)\phi_s(v,x)$ *for* $s = y$ *or* z.

Theorem 2 shows that the $\widehat{SM}$ test has nontrivial power against $\mathbb{H}_1(a_n)$ with $a_n = n^{-1/2}h^{-d_x/4}$. In terms of Pitman's criterion, it is asymptotically more efficient than the nonparametric tests of Su and White (2007, 2008, 2014), Huang (2010), Bouezmarni et al. (2012), Bouezmarni and Taamouti (2014), and Taamouti et al. (2014). This is because $\widehat{SM}$ only involves d_x-dimensional smoothing, whereas the aforementioned tests involve smoothing of $d_x + d_y + d_z$ or at least $d_x + d_z$ dimensions. Thus, they can only detect local alternatives with a rate of $n^{-1/2}h^{-(d_x+d_y+d_z)/4}$ or $n^{-1/2}h^{-(d_x+d_z)/4}$. On the other hand, Huang's (2010) test can detect local alternatives with a rate of $n^{-1/2}h^{-d_x/2}n_X^{-1/4}(p_nq_n)^{1/2}$, where $p_n \to \infty$ and $q_n \to \infty$ are the maximum orders of nonparametric series approximations, and $n_X \to \infty$ is the number of grid points of X. This rate is slower than $n^{-1/2}h^{-d_x/4}$ since n_X, p_n, q_n can grow only as a power function of $\ln n$.

It should be noted that the tests of Su and White (2012) and Huang et al. (2016) can detect $\mathbb{H}_1(a_n)$ with $a_n = n^{-1/2}$, which is faster than $a_n = n^{-1/2}h^{-d_x/4}$ for the

$\widehat{SM}$ test. However, this conclusion is peculiar to the class of smooth type local alternatives in (16). Suppose we consider

$$\mathbb{H}_2(a_n,b_n): \quad \phi_{yz}(u,v,x) = \phi_y(u,x)\phi_z(v,x) + a_n\delta\left(u,v,\frac{x-c}{b_n}\right),$$

where, given each pair (u,v), $\delta(u,v,\cdot)$ is bounded and continuously differentiable up to order $p+1$ in the interior of $\mathbb{G}$, c is a constant, $a_n \to 0$, $b_n \to 0$ as $n \to \infty$, $a_n^2 b_n = n^{-1}h^{-d_x/2}$, and $h = o(b_n)$. This type of local alternative has been considered by Rosenblatt (1975), Horowitz and Spokoiny (2001) and Su and White (2008) among others. It can arise (e.g.,) when

$$f(y,z|x) = f(y|x)f(z|x) + a_n q(y,z|x),$$

where $q(y,z|x) = q_0(y,z)l\left(\frac{x-c}{b_n}\right)$, $q_0(y,z)$ is bounded and square-integrable with $q_0(y,z) \neq 0$, $\iint q_0(y,z)dydz = 0$, and $l(\cdot)$ is bounded and continuously differentiable up to order $p+1$, possibly with unbounded support. Under this kind of local alternative, the deviation of $\mathbb{H}_2(a_n,b_n)$ from $\mathbb{H}_0$ has a nonsmooth spike at location c. That is, Y_t and Z_t display strong mutual dependence when X_t takes values in a neighborhood of point c but little elsewhere. The shrinkage parameter b_n measures the effective size of the neighborhood of point c, and a_n controls the speed at which the deviation of $\mathbb{H}_2(a_n,b_n)$ from $\mathbb{H}_0$ vanishes to 0. It is not difficult to see that the departure of $\mathbb{H}_2(a_n,b_n)$ from $\mathbb{H}_0$ is of order a_n when X_t takes values in the neighborhood of c with size b_n, but is of a higher order for any other distinct point of X_t on the compact set $\mathbb{G}$. See DGP.P3 for a time series example in Section 7.

We now derive the asymptotic power of the $\widehat{SM}$ test under $\mathbb{H}_2(a_n,b_n)$.

THEOREM 3. *Suppose Assumptions A.1–A.4 hold and the bandwidth* $h = cn^{-\lambda}$ *for* $\frac{2}{d_x+4p+4} < \lambda < \frac{2}{3d_x}$, *where* $0 < c < \infty$. *Then, under* $\mathbb{H}_2(a_n,b_n)$ *with* $a_n \to 0$, $b_n \to 0$, $a_n^2 b_n = n^{-1}h^{-d_x/2}$, *and* $h = o(b_n)$, *we have* $P\left[\widehat{SM} \geq z_\alpha | \mathbb{H}_2(a_n,b_n)\right] \to 1 - \Phi(z_\alpha - \kappa/\sqrt{V})$ *as* $n \to \infty$, *where* $\kappa = a(c)g(c) \iiint |\delta(u,v,w)|^2 dW_1(u)dW_2(v)dw$.

Theorem 3 implies that $\widehat{SM}$ has nontrivial power under $\mathbb{H}_2(a_n,b_n)$ with $a_n b_n^{1/2} = n^{-1/2}h^{-d_x/4}$. This is because under $\mathbb{H}_2(a_n,b_n)$, the noncentrality parameter of $\widehat{SM}$ depends on the squared departure $\iiint |\sigma(u,v,x)|^2 dW_1(u)dW_2(v)a(x)g(x)dx = a_n^2 b_n a(c)g(c) \iiint |\delta(u,v,w)|^2 dW_1(u)dW_2(v)dw$. The $\widehat{SM}$ test is asymptotically more efficient than the tests of Su and White (2007, 2008, 2012, 2014), Huang (2010), Bouezmarni et al. (2012), Bouezmarni and Taamouti (2014), and Taamouti et al. (2014) under $\mathbb{H}_2(a_n,b_n)$. The latter cannot detect the rate $a_n b_n^{1/2} = n^{-1/2}h^{-d_x/4}$, due to additional smoothing of other variables. For Su and White's (2012) test, a careful inspection shows that its noncentrality parameter depends on the integral $\int \sigma(u,v,x)\psi(x)dx = a_n \int \delta\left(u,v,\frac{x-c}{b_n}\right)\psi(x)dx = a_n b_n \psi(c) \int \delta(u,v,w)dw$ for some smooth weighting function $\psi(x)$. The rate $a_n b_n$ can be faster than $n^{-1/2}$, rendering Su and White's (2012) test unable to detect $\mathbb{H}_2(a_n, b_n)$. For example, if $d_x = 1, h = n^{-1/2}, b_n = h^{5/6}$ and $a_n = n^{-1/2}h^{-2/3}$, then $a_n \int \delta\big(u, v, \frac{x-c}{b_n}\big)\psi(x)dx$ is of order $a_n b_n = n^{-7/12}$, which is faster than $n^{-1/2}$. Hence, Su and White's (2012) test fails to detect $\mathbb{H}_2(a_n, b_n)$ in this example. Similarly, although Huang et al.'s (2016) test can detect the rate $n^{-1/2}$ for $\mathbb{H}_1(a_n)$, it may fail to detect $\mathbb{H}_2(a_n, b_n)$.

6. INFERENCE ON PATTERNS OF CONDITIONAL DEPENDENCE

When $\mathbb{H}_0$ is rejected, one may like to gauge possible reasons of rejection, which can provide valuable information for modeling economic relationships. For example, if we know that two variables have conditional dependence in mean, then we can use a conditional mean model.

As is well known, the characteristic function can be differentiated to obtain various moments (when these exist). As the omnibus test $\widehat{SM}$ is based on the conditional characteristic function, we can develop a class of derivative tests to capture various patterns of conditional dependence. The derivative tests can check various hypotheses of interest, including omitted variables, Granger causality in various moments, and conditional uncorrelatedness. As an important feature, these derivative tests are all model-free.

6.1. Inference on Conditional Dependence of Various Moments

Suppose the m-th order moment of Y_t exists. For $\sigma(u, v, x)$ in (3), taking the m-th order partial derivative with respect to u at $u = 0$, we obtain

$$\sigma^{(m)}(0, v, x) = \frac{\partial^m \sigma(u, v, x)}{\partial u^m}|_{u=0} = \mathbf{i}^m \mathrm{cov}\big(Y_t^m, e^{\mathbf{i}v'Z_t}|X_t = x\big), \tag{18}$$

for any $m = 1, 2, \ldots$. Under the null hypothesis

$$\mathbb{H}_0^{(m)} : P\big[\mathrm{cov}\big(Y_t^m, e^{\mathbf{i}v'Z_t}|X_t\big) = 0\big] = 1 \text{ for all } v \in \mathbb{R}^{d_z}, \tag{19}$$

we have $\sigma^{(m)}(0, v, x) = 0$ for all $v \in \mathbb{R}^{d_z}$ and all $x \in \mathbb{G}$. Thus, we can test $\mathbb{H}_0^{(m)}$ in (19) by examining whether $\sigma^{(m)}(0, v, x) = 0$. Denote the estimator of $\sigma^{(m)}(0, v, x)$ as $\hat{\sigma}^{(m)}(0, v, x) = \frac{\partial^m}{\partial u^m}\hat{\sigma}(u, v, x)|_{u=0}$. Based on the notation of $\hat{\sigma}(u, v, x) = \hat{\phi}_{yz}(u, v, x) - \hat{\phi}_y(u, x)\hat{\phi}_z(v, x)$, we have

$$\hat{\sigma}^{(m)}(0, v, x) = \frac{\partial^m}{\partial u^m}\hat{\phi}_{yz}(u, v, x)|_{u=0} - \frac{\partial^m}{\partial u^m}\hat{\phi}_y(u, x)|_{u=0}\hat{\phi}_z(v, x),$$

where

$$\frac{\partial^m}{\partial u^m}\hat{\phi}_{yz}(u, v, x)|_{u=0} = \frac{1}{nh^{d_x}g(x)}\sum_{t=1}^n D\left(p, \frac{X_t - x}{h}\right) K\left(\frac{X_t - x}{h}\right)\mathbf{i}^m Y_t^m e^{\mathbf{i}v'Z_t}[1 + o_P(1)]$$

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

and

$$\frac{\partial^m}{\partial u^m}\hat{\phi}_y(u,x)|_{u=0} = \frac{1}{nh^{d_x}g(x)}\sum_{t=1}^{n} D\left(p,\frac{X_t-x}{h}\right)K\left(\frac{X_t-x}{h}\right)\mathbf{i}^m Y_t^m[1+o_P(1)].$$

Similar to the construction of $\widehat{SM}$, we use the following quadratic form to test $\mathbb{H}_0^{(m)}$ in (19):

$$\hat{M}^{(m)} = \frac{1}{n}\sum_{t=1}^{n}\int \left|\hat{\sigma}^{(m)}(0,v,X_t)\right|^2 a(X_t)dW_2(v).$$

Following the proof of Theorem 1, we can show that under $\mathbb{H}_0^{(m)}$ in (19) and other regularity conditions,

$$\widehat{SM}^{(m)} = \left[nh^{d_x/2}\hat{M}^{(m)} - \hat{B}^{(m)}\right]/\sqrt{\hat{V}^{(m)}} \xrightarrow{d} N(0,1),$$

where

$$\hat{B}^{(m)} = h^{-dx/2}\iint \left[\hat{\phi}_y^{(2m)}(0,x) - \left|\hat{\phi}_y^{(m)}(0,x)\right|^2\right]\left[1-|\hat{\phi}_z(v,x)|^2\right]dW_2(v)a(x)dx$$
$$\times \int D^2(p,\tau)K^2(\tau)d\tau,$$
$$\hat{V}^{(m)} = 2\iiint \left[\hat{\phi}_y^{(2m)}(0,x) - |\hat{\phi}_y^{(m)}(0,x)|^2\right]^2\left|\hat{\Phi}_z(v_1+v_2,x)\right|^2 dW_2(v_1)dW_2(v_2)a^2(x)dx$$
$$\times \int\left[\int D(p,\tau)D(p,\tau+\eta)K(\tau)K(\tau+\eta)d\tau\right]^2 d\eta,$$

with $\hat{\phi}_y^{(s)}(0,x) = \frac{\partial^s}{\partial u^s}\hat{\phi}_y(u,x)|_{u=0}$.

Moreover, to improve the size of the test in finite samples, we can use a finite sample version of $\widehat{SM}^{(m)}$, denoted as $\widehat{SM}_n^{(m)}$, by replacing $\hat{B}^{(m)}$ with its finite sample version

$$\hat{B}_n^{(m)} = h^{d_x/2}\sum_{t=1}^{n}\sum_{s=1}^{n} a(X_t)\hat{W}\left(\frac{X_s-X_t}{h},X_t\right)^2\left[Y_s^m - \hat{\phi}_y^{(m)}(0,X_s)\right]^2\int\left|\hat{\varepsilon}_z(v,X_s)\right|^2 dW_2(v).$$

We now discuss the primary case of $p=1$. The $\widehat{SM}^{(1)}$ test checks whether $\text{cov}(Y_t, e^{iv'Z_t}|X_t) = 0$, which is equivalent to the model-free hypothesis of $E(Y_t|X_t,Z_t) = E(Y_t|X_t)$, i.e., Z_t is not an omitted variable. In a cross-sectional context, Fan and Li (1996) and Lavergne and Vuong (2000) develop nonparametric tests for omitted variables using a weighted average of squared conditional mean estimates of residuals. Aït-Sahalia et al. (2001) also consider a nonparametric test for omitted variables in a time series context. As $\widehat{SM}^{(1)}$ only involves d_x-dimensional smoothing, it is more powerful than the tests of Fan and Li (1996), Lavergne and Vuong (2000) and Aït-Sahalia et al. (2001). The latter all involve (d_x+d_z)-dimensional smoothing and are therefore asymptotically less efficient.

The $\widehat{SM}^{(m)}$ test can be applied to check Granger causality in mth moment. Put $X_t = Y_{t-p}^{t-1} = \left(Y_{t-1},\ldots,Y_{t-p}\right)'$ and $Z_t = X_{t-q}^{t-1} = \left(X_{t-1},\ldots,X_{t-q}\right)'$ for lag orders p,q. Then the null hypothesis becomes $\mathbb{H}_0^{(m)}: E(Y_t^m|Y_{t-p}^{t-1}, X_{t-q}^{t-1}) = E(Y_t^m|Y_{t-p}^{t-1})$, i.e., there is no Granger causality in the mth moment of Y_t from X_{t-q}^{t-1}. The choice of $m=1$ delivers a test for Granger causality in mean. Compared with Granger's (1969) F test, $\widehat{SM}^{(1)}$ is model-free, and it is powerful in capturing not only linear but also various nonlinear relationships in mean, including ARCH-in-mean effects (Engle, Lilien, and Robins, 1987), threshold effects (Tong and Lim, 1980), and functional coefficient autoregressive effects (Priestley, 1988; Chen and Tsay, 1993). Nishiyama, Hitomi, Kawasaki, and Jeong (2011) also propose a model-free test for Granger causality in mean and high order moments. Their test could achieve the parametric rate $a_n = n^{-1/2}$ for a class of smooth local alternatives $\mathbb{H}_1^{(m)}(a_n): Y_t^m = g(Y_{t-p}^{t-1}) + a_n\kappa(Y_{t-p}^{t-1}, X_{t-q}^{t-1})$, where $g(\cdot)$ and $\kappa(\cdot)$ are smooth functions. However, under a class of nonsmooth local alternatives $\mathbb{H}_2^{(m)}(a_n,b_n): Y_t^m = g(Y_{t-p}^{t-1}) + a_n\kappa\left(X_{t-q}^{t-1}\right)l\left(\frac{Y_{t-p}^{t-1}-c}{b_n}\right)$, where $\kappa(\cdot)$ and $l(\cdot)$ satisfy certain regularity conditions, $\widehat{SM}^{(m)}$ could be asymptotically more powerful than Nishiyama et al.'s (2011) test. In addition, unlike Nishiyama et al.'s (2011) test, $\widehat{SM}^{(m)}$ has a convenient asymptotic null $N(0,1)$ distribution. Our simulations below show that $\widehat{SM}^{(1)}$ outperforms Nishiyama et al.'s (2011) test in finite samples.

6.2. Inference on Conditional Correlation Between Moments

Suppose the m-th and l-th order moments of Y_t and Z_t exist respectively. Then taking the m-th and l-th order partial derivative of $\sigma(u,v,x)$ with respect to (u,v) at the origin, we obtain

$$\sigma^{(m,l)}(0,0,x) = \frac{\partial^{m+l}\sigma(u,v,x)}{\partial u^m \partial v^l}|_{(u,v)=(0,0)} = \mathbf{i}^{m+l}\mathrm{cov}\left(Y_t^m, Z_t^l \middle| X_t = x\right) \tag{20}$$

for any $m = 1,2,\ldots; l = 1,2,\ldots$. Under the null hypothesis

$$\mathbb{H}_0^{(m,l)}: P\left[\mathrm{cov}\left(Y_t^m, Z_t^l \middle| X_t\right) = 0\right] = 1, \tag{21}$$

we have $\sigma^{(m,l)}(0,0,x) = 0$ for all $x \in \mathbb{G}$. Like in Section 6.1, we have

$$\hat{\sigma}^{(m,l)}(0,0,x) = \frac{\partial^{m+l}}{\partial u^m \partial v^l}\hat{\phi}_{yz}(u,v,x)|_{(u,v)=(0,0)} - \frac{\partial^m}{\partial u^m}\hat{\phi}_y(u,x)|_{u=0}\frac{\partial^l}{\partial v^l}\hat{\phi}_z(v,x)|_{v=0},$$

where

$$\frac{\partial^{m+l}}{\partial u^m \partial v^l}\hat{\phi}_{yz}(u,v,x)|_{(u,v)=(0,0)} = \frac{1}{nh^{d_x}g(x)}\sum_{t=1}^n D\left(p, \frac{X_t - x}{h}\right)K\left(\frac{X_t - x}{h}\right)\mathbf{i}^{m+l}Y_t^m Z_t^l[1 + o_P(1)]$$

and

$$\frac{\partial^l}{\partial v^l}\hat{\phi}_z(v,x)|_{v=0} = \frac{1}{nh^{d_x}g(x)}\sum_{t=1}^n D\left(p, \frac{X_t - x}{h}\right)K\left(\frac{X_t - x}{h}\right)\mathbf{i}^l Z_t^l[1 + o_P(1)].$$

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

Then we can use the statistic

$$\hat{M}^{(m,l)} = \frac{1}{n}\sum_{t=1}^{n} a(X_t)\left|\hat{\sigma}^{(m,l)}(0,0,X_t)\right|^2 \tag{22}$$

to check conditional uncorrelatedness between Y_t^m and Z_t^l given X_t. We could prove that under $\mathbb{H}_0^{(m,l)}$ and suitable regularity conditions, a standardized version of $\hat{M}^{(m,l)}$, namely

$$\widehat{SM}^{(m,l)} = \left[nh^{d_x/2}\hat{M}^{(m,l)} - \hat{B}^{(m,l)}\right]/\sqrt{\hat{V}^{(m,l)}} \xrightarrow{d} N(0,1),$$

where

$$\hat{B}^{(m,l)} = h^{-dx/2}\int \left[\hat{\phi}_y^{(2m)}(0,x) - \left|\hat{\phi}_y^{(m)}(0,x)\right|^2\right]\left[\hat{\phi}_z^{(2l)}(0,x) - \left|\hat{\phi}_z^{(l)}(0,x)\right|^2\right] a(x)dx \times \int D^2(p,\tau)K^2(\tau)d\tau,$$

$$\hat{V}^{(m,l)} = 2\int \left[\hat{\phi}_y^{(2m)}(0,x) - \left|\hat{\phi}_y^{(m)}(0,x)\right|^2\right]^2\left[\hat{\phi}_z^{(2l)}(0,x) - \left|\hat{\phi}_z^{(l)}(0,x)\right|^2\right]^2 a^2(x)dx \times \int \left[\int D(p,\tau)D(p,\tau+\eta)K(\tau)K(\tau+\eta)d\tau\right]^2 d\eta.$$

Again, to improve the size of the test in finite samples, we can use a finite sample version of $\widehat{SM}^{(m,l)}$, denoted as $\widehat{SM}_n^{(m,l)}$, by replacing $\hat{B}^{(m,l)}$ and $\hat{V}^{(m,l)}$ respectively with

$$\hat{B}_n^{(m,l)} = h^{d_x/2}\sum_{t=1}^{n}\sum_{s=1}^{n} a(X_t)\hat{W}\left(\frac{X_s - X_t}{h}, X_t\right)^2 \hat{e}_y^{(m)}(0,X_s)^2\hat{e}_z^{(l)}(0,X_s)^2,$$

$$\hat{V}_n^{(m,l)} = 2h^{d_x/2}\sum_{1\le r<s\le n}\left[\sum_{t=1}^{n} a(X_t)\hat{W}\left(\frac{X_s - X_t}{h}, X_t\right)\hat{W}\left(\frac{X_r - X_t}{h}, X_t\right) \times \hat{e}_y^{(m)}(0,X_s)\hat{e}_z^{(l)}(0,X_s)\hat{e}_y^{(m)}(0,X_r)\hat{e}_z^{(l)}(0,X_r)\right]^2,$$

where $\hat{e}_y^{(m)}(0,X_s) = Y_s^m - \hat{\phi}_y^{(m)}(0,X_s)$ and $e_z^{(l)}(0,X_s) = Z_s^l - \hat{\phi}_z^{(l)}(0,X_s)$.

The choice of derivative orders (m,l) allows us to examine various conditional correlations between the powers of Y_t and Z_t. For example, the choice of $(m,l) = (1,1)$ yields a model-free test for conditional uncorrelatedness (i.e., $\text{cov}(Y_t, Z_t|X_t) = 0$). For a time series $\{Y_t\}$ and a lag order $k \ge 2$, we put $Z_t = Y_{t-k}$ and $X_t = Y_{t-k+1}^{t-1} = (Y_{t-1},\ldots,Y_{t-k+1})$. Then $\gamma(k, y_{t-k+1}^{t-1}) = \text{cov}(Y_t, Y_{t-k}|Y_{t-k+1}^{t-1} = y_{t-k+1}^{t-1})$ is the well-known PACF. It follows that $\widehat{SM}^{(1,1)}$ is a weighted average of squared PACFs and could be used to test whether a higher lag order conditional on lower lag orders is significant in a nonparametric autoregression. Compared with the commonly used t statistic, $\widehat{SM}^{(1,1)}$ not only

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

avoids the misspecification problem, but is also powerful in detecting nonlinear dependence.

7. MONTE CARLO STUDY

We now study the finite sample performance of the $\widehat{SM}$ test for Granger causality in a time series context. We compare $\widehat{SM}$ with the tests of Granger (1969), Su and White (2007, 2012) and Nishiyama et al. (2011). For the derivative tests $\widehat{SM}^{(m)}$ and $\widehat{SM}^{(m,l)}$, we consider the primitive cases of $m = 1$ and $(m,l) = (1,1)$. Here, $\widehat{SM}^{(1)}$ checks whether Z_t is an omitted variable in modeling $E(Y_t|X_t)$, and $\widehat{SM}^{(1,1)}$ checks conditional uncorrelatedness between Y_t and Z_t given X_t.

To examine the size and power of $\widehat{SM}, \widehat{SM}^{(1)}, \widehat{SM}^{(1,1)}$, we consider the following DGPs:

DGP.S1: $Y_t = 0.5Y_{t-1} + \varepsilon_{1,t}$;

DGP.S2: $Y_t = \sqrt{h_t}\varepsilon_{1,t}, h_t = 0.01 + 0.5Y_{t-1}^2$;

DGP.S3: $Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}, h_{1,t} = 0.01 + 0.9h_{1,t-1} + 0.05Y_{t-1}^2,$
$Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}, h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2$;

DGP.P1: $Y_t = 0.5Y_{t-1} + 0.5Z_{t-1} + \varepsilon_{1,t}$;

DGP.P2: $Y_t = 0.5Y_{t-1}Z_{t-1} + \varepsilon_{1,t}$;

DGP.P3: $Y_t = 0.5Y_{t-1} + 4\varphi(Y_{t-1}/0.1)Z_{t-1} + \varepsilon_{1,t},\ \varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$;

DGP.P4: $Y_t = 0.4Y_{t-1} + 0.2Z_{t-1}^2 + \varepsilon_{1,t}$;

DGP.P5: $Y_t = 0.3 + 0.2\log(h_t) + \sqrt{h_t}\varepsilon_{1,t}, h_t = 0.01 + 0.5Y_{t-1}^2 + 0.3Z_{t-1}^2$;

DGP.P6: $Y_t = 0.5Y_{t-1} + 0.5Z_{t-1}\varepsilon_{1,t}$;

DGP.P7: $Y_t = \sqrt{h_t}\varepsilon_{1,t}, h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Z_{t-1}^2$;

DGP.P8: $Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}, h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4Y_{t-1}^2 + 0.5Z_{t-1}^2,$
$Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}, h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2.$

Here $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$ except in DGP.S3 and DGP.P8, and $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{2,t}\}$ are two mutually independent i.i.d.$N(0,1)$ sequences. All DGPs except DGP.P3 and DGP.P5 are investigated by Su and White (2008). DGP.P3 is an example of the class of nonsmooth local alternatives $\mathbb{H}_2(a_n, b_n)$ considered in Section 5, where Y_t depends on Z_{t-1} when and only when Y_{t-1} takes values in a narrow interval of 0, and DGP.P5 is an ARCH-in-mean process proposed by Engle et al. (1987). These DGPs cover a wide range of linear and nonlinear time series. Here, we test whether Y_t is independent of Z_{t-1} conditional on Y_{t-1}, that is, whether Z_t Granger-causes Y_t at lag order 1. We use DGP.S1–S3 to examine the sizes of the tests and DGP.P1–P8 to examine the powers. All DGPs except DGP.P1 have nonlinear dependence in mean or variance or both. Under DGP.P4–P8, the null hypothesis of conditional uncorrelatedness holds, and under DGP.P6–P8, the null hypothesis of no Granger causality in mean holds.

For $\widehat{SM}$, $\widehat{SM}^{(1)}$ and $\widehat{SM}^{(1,1)}$, following Aït-Sahalia et al. (2001) and Chen and Hong (2010), we choose the Gaussian kernel and $a(X_t) = \mathbf{1}(|X_t| \leq 1.5)$, where $\mathbf{1}(\cdot)$ is the indicator function and X_t is standardized by its sample mean and standard deviation. We have also tried the Gaussian weighting function for $a(\cdot)$ and the results are similar to the truncated weighting function. We choose the $N(0,1)$ CDF for both $W_1(\cdot)$ and $W_2(\cdot)$ and choose the bandwidth $h = n^{-4/17}$.

For the tests of Su and White (2012) and Nishiyama et al. (2011), we follow their simulation designs. For Su and White's (2007) test, we follow Su and White (2007) to choose the fourth order kernel $k(u) = \frac{1}{2\sqrt{2\pi}}(3-u^2)e^{-u^2/2}$. To make Su and White's (2007) test and ours comparable, we choose $h_1 = n^{-4/17}$ and $h_2 = n^{-1/3}$, which satisfy Assumption $A.2$ in Su and White (2007). Since smoothed nonparametric tests are usually sensitive to the choice of bandwidth, we use the local bootstrap proposed by Paparoditis and Politis (2000) and modified by Su and White (2008). Conditional on a sample $\{X_t, Y_t, Z_t\}_{t=1}^n$, we draw a bootstrap sample $\{X_t^*, Y_t^*, Z_t^*\}_{t=1}^n$ as follows: (i) draw $\{X_t^*\}$ from the smoothed kernel density $\tilde{f}_b(x) = n^{-1}\sum_{t=1}^n K_{b,h_b}(X_t - x)$, where $K_{b,h_b}(x) = h_b^{-d_x} K_b(x/h_b)$ with $K_b(\cdot)$ being a product of a univariate bootstrap kernel $k_b(\cdot)$ and a resampling bandwidth h_b; (ii) given $X_t^*,\ t = 1,\ldots,n$, draw Y_t^* and Z_t^* independently from the smoothed conditional densities $\tilde{f}(y|X_t^*) = \sum_{s=1}^n K_{b,h_b}(Y_s - y) K_{b,h_b}(X_s - X_t^*)/\sum_{r=1}^n K_{b,h_b}(X_r - X_t^*)$ and $\tilde{f}(z|X_t^*) = \sum_{s=1}^n K_{b,h_b}(Z_s - z) K_{b,h_b}(X_s - X_t^*)/\sum_{r=1}^n K_{b,h_b}(X_r - X_t^*)$; (iii) repeat steps (i) and (ii) B times given each sample $\{X_t, Y_t, Z_t\}_{t=1}^n$. The argument for the validity of the local bootstrap here is the same as that of Su and White (2008) and hence is omitted. For each DGP, we generate 500 data sets with $n = 100, 200$ respectively and we set $B = 100$. We use the Gaussian kernel and $h_b = n^{-4/17}$ as the bootstrap kernel and resampling bandwidth.

Table 1 reports the empirical sizes of various tests under DGP.S1–S3 at both 10% and 5% significance levels, using the bootstrap. The tests $\widehat{SM}$, $\widehat{SM}^{(1)}$, and $\widehat{SM}^{(1,1)}$ have reasonable sizes. The tests of Granger (1969), Su and White (2007, 2012) and Nishiyama et al. (2011) also have good sizes.

We also consider the results based on asymptotic critical values. For each DGP, we simulate 1,000 data sets with $n = 100, 200, 500, 1,000$ respectively. The results for our tests, which are reported in the online Supplementary Material, are similar to those based on bootstrap critical values, especially for the size performance. This confirms the advantage of allowing the nonparametric estimators of both conditional joint and marginal characteristic functions to jointly determine the asymptotic distributions of $\widehat{SM}$, $\widehat{SM}^{(1)}$, and $\widehat{SM}^{(1,1)}$. Since these tests have achieved reasonable sizes using asymptotic approximation, it is not surprising to observe the inappreciable role of bootstrap approximation. In contrast, the tests of Su and White (2007) and Nishiyama et al. (2011) suffer from severe

TABLE 1. Size of tests under DGP.S1–S3

		*SW*07		*SW*12		$\widehat{SM}$		*NHKJ*11		$\widehat{SM}^{(1)}$		$\widehat{SM}^{(1,1)}$		*LIN*	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.S1	$n=100$	0.056	0.098	0.048	0.106	0.044	0.082	0.056	0.108	0.034	0.078	0.048	0.088	0.050	0.106
	$n=200$	0.050	0.076	0.056	0.094	0.040	0.096	0.060	0.106	0.044	0.094	0.046	0.092	0.052	0.102
DGP.S2	$n=100$	0.062	0.116	0.048	0.108	0.064	0.106	0.052	0.114	0.064	0.118	0.052	0.104	0.047	0.103
	$n=200$	0.066	0.122	0.052	0.100	0.076	0.116	0.074	0.128	0.072	0.134	0.050	0.120	0.057	0.108
DGP.S3	$n=100$	0.048	0.080	0.038	0.094	0.048	0.106	0.064	0.120	0.052	0.120	0.074	0.120	0.042	0.094
	$n=200$	0.048	0.094	0.048	0.096	0.064	0.098	0.062	0.114	0.062	0.108	0.064	0.108	0.046	0.102

Notes: (i) *SW*07 and *SW*12 denote the tests of Su and White (2007) and Su and White (2012), *NHKJ*11 denotes Nishiyama et al.'s (2011) test, and *LIN* denotes Granger's (1969) *F* test; (ii) the results for all tests except *LIN* are based on bootstrap critical values.

overrejection/underrejection when using asymptotic critical values, but they have remarkable improvement using the bootstrap. Moreover, we examine the performance of all the tests using $h = cn^{-4/17}$ with $c = 0.5, 1.5$. The results are similar to those with $c = 1$. To examine the effect of increasing the dimension of conditioning variables, we also consider testing whether $Y_t \perp Z_{t-1} | (Y_{t-1}, Y_{t-2})$. The results show that $\widehat{SM}$ tends to overreject under $\mathbb{H}_0$ when n is small, but the size improves as n increases. The results with $c \neq 1$, $d_x \neq 1$ and the use of the Gaussian weighting function $a(\cdot)$ are reported in the online Supplementary Material.

We now turn to examine power. Table 2 reports the rejection rates of various tests under DGP.P1–P8 at the 10% and 5% levels, using bootstrap critical values. We have also used the empirical critical values obtained under $\mathbb{H}_0$ to compute size-corrected power so as to compare all tests on an equal ground. The results on the relative performance among the tests, reported in the online Supplementary Material, are similar to those based on bootstrap critical values. Table 2 shows that Granger's (1969) *F* test is most powerful under DGP.P1, which has a linear Granger causality relationship. In DGP.P2–P3, there exist linear relationships between Y_t and Z_{t-1} conditional on Y_{t-1}. Interestingly, Granger's (1969) *F* test has little power against DGP.P2 but good power against DGP.P3. This is because under DGP.P2, the weights with which Y_{t-1} takes values symmetric about 0 average to 0. In contrast, under DGP.P3, the weight $\varphi(Y_{t-1}/0.1)$ is always positive no matter whether Y_{t-1} is positive or negative.

The $\widehat{SM}$ test is powerful in detecting all DGP.P1–P8, and is generally more powerful than Su and White's (2007) test. This is consistent with our analysis on the relative efficiency between our test and Su and White's (2007) test. In addition, although Su and White's (2012) test could achieve the parametric convergence rate under $\mathbb{H}_1(a_n)$, it is less powerful than $\widehat{SM}$ against several DGPs, especially DGP.P3, which is a nonsmooth local alternative under $\mathbb{H}_2(a_n, b_n)$. It is interesting to observe that $\widehat{SM}^{(1)}$ is powerful in capturing various forms of Granger causality in mean under DGP.P1–P5 and is robust to higher order conditional dependence such as ARCH effects under DGP.P6–P8. Moreover, $\widehat{SM}^{(1)}$ is more powerful than Nishiyama et al.'s (2011) test. Similarly, $\widehat{SM}^{(1,1)}$ is powerful in capturing various

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

TABLE 2. Power of tests under DGP.P1–P8

		$SW07$		$SW12$		$\widehat{SM}$		$NHKJ11$		$\widehat{SM}^{(1)}$		$\widehat{SM}^{(1,1)}$		LIN	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.P1	$n = 100$	0.428	0.550	0.956	0.992	0.860	0.920	0.266	0.362	0.938	0.966	0.924	0.976	1.00	1.00
	$n = 200$	0.786	0.850	1.00	1.00	0.990	0.996	0.332	0.424	0.998	1.00	0.994	1.00	1.00	1.00
DGP.P2	$n = 100$	0.362	0.510	0.566	0.764	0.778	0.852	0.194	0.296	0.902	0.938	0.924	0.970	0.206	0.290
	$n = 200$	0.676	0.762	0.952	0.980	0.974	0.988	0.222	0.322	0.992	0.996	0.994	1.00	0.215	0.286
DGP.P3	$n = 100$	0.090	0.150	0.154	0.254	0.190	0.264	0.072	0.136	0.238	0.384	0.236	0.358	0.260	0.357
	$n = 200$	0.188	0.290	0.268	0.420	0.384	0.538	0.090	0.154	0.576	0.688	0.578	0.720	0.500	0.617
DGP.P4	$n = 100$	0.130	0.212	0.376	0.562	0.234	0.346	0.078	0.130	0.402	0.540	0.050	0.112	0.175	0.267
	$n = 200$	0.152	0.256	0.774	0.872	0.512	0.632	0.064	0.126	0.712	0.836	0.042	0.086	0.149	0.219
DGP.P5	$n = 100$	0.552	0.682	0.624	0.798	0.578	0.720	0.122	0.222	0.234	0.372	0.072	0.144	0.191	0.264
	$n = 200$	0.844	0.922	0.952	0.992	0.898	0.960	0.142	0.244	0.372	0.510	0.072	0.120	0.149	0.227
DGP.P6	$n = 100$	0.674	0.802	0.412	0.700	0.882	0.946	0.116	0.186	0.110	0.202	0.044	0.088	0.250	0.344
	$n = 200$	0.944	0.982	0.936	0.994	0.998	1.00	0.090	0.156	0.120	0.174	0.020	0.052	0.223	0.310
DGP.P7	$n = 100$	0.366	0.504	0.244	0.514	0.476	0.648	0.054	0.100	0.062	0.118	0.034	0.088	0.163	0.243
	$n = 200$	0.598	0.740	0.640	0.870	0.806	0.914	0.068	0.122	0.078	0.148	0.058	0.114	0.147	0.227
DGP.P8	$n = 100$	0.278	0.384	0.134	0.272	0.344	0.494	0.050	0.112	0.090	0.148	0.042	0.094	0.175	0.268
	$n = 200$	0.404	0.560	0.334	0.582	0.686	0.830	0.084	0.136	0.080	0.154	0.050	0.102	0.215	0.167

Notes: See the notes in Table 1.

forms of conditional correlation between Y_t and Z_{t-1}, and is robust to serial dependence in higher order moments. Indeed, under DGP.P4–P8, for which there exists no conditional correlation between Y_t and Z_{t-1} but there exists serial dependence in higher order moments, the rejection rates of $\widehat{SM}^{(1,1)}$ are close to the nominal significance levels, implying that $\widehat{SM}^{(1,1)}$ has robust reasonable sizes. Finally, we note that $\widehat{SM}^{(1,1)}$ is powerful in capturing some nonlinear Granger relationships in mean such as DGP.P2, whereas Granger's F test is silent about this DGP.

8. APPLICATION TO GRANGER CAUSALITY BETWEEN MONEY AND OUTPUT

The relationship between money and output has attracted substantial sustained interest from macroeconomists and policy makers. This issue not only reflects the causal relationship between nominal economic variables (e.g., money) and real economic variables (e.g., output), but also involves the important problem about whether monetary policy is neutral. Many studies have investigated the relationships between output and money, including Sims (1972, 1980), Christiano and Ljungqvist (1988), Stock and Watson (1989), and Friedman and Kuttner (1993). The results vary with different sample periods. In fact, a stream of economic theories imply a nonlinear relationship between money and output. The sources of nonlinear effects between money and output may include nonlinear wage indexation and price adjustment (Kandil, 1995), asymmetric preference of the central bank's monetary policy (Nobay and Peel, 2003), nonlinearity of aggregate supply and demand curves, and so on. However, most empirical studies have employed linear Granger causality tests, which may have little power for nonlinear relationships.

In this section, we will use our tests to study various Granger causalities between money and output. We use U.S. monthly data in the period 1959:M1-2012:M6, with 642 observations. We measure output by monthly Industrial Production Index (IPI). Following Psaradakis, Ravn, and Sola (2005), we use three monetary variables—the narrow money supply M1, the broad money supply M2 and the Federal Funds rate (ir), as the proxies for the U.S. monetary policy. We logarithmically transform IPI, M1, M2, denoted as ipi, $m1$, $m2$ respectively. All data except the interest rate are seasonally adjusted. We check stationarity of the data by the augmented Dickey–Fuller test. The results suggest that ipi, $m1$, $m2$, ir are integrated of order 1, and the differenced series, denoted as Δipi, $\Delta m1$, $\Delta m2$, Δir, are weakly stationary. Since the Federal Reserve Board adjusts its target interest rate by a multiple of 25 basis points, not by a percentage of the interest level, it is more appropriate to assume its difference rather than its logarithmic difference to be stationary (Bae and de Jone, 2007). We test various Granger causalities among these differenced series. The plots of these differenced series are given in the online Supplementary Material.

Granger's (1969) F test checks whether output (Δipi) and money $(\Delta m1$, $\Delta m2$, $\Delta ir)$ Granger-cause each other in the following linear regressions:

$$\Delta ipi_t = \alpha_0 + \alpha_1 \Delta ipi_{t-1} + \cdots + \alpha_p \Delta ipi_{t-p} + \beta_1 \Delta m_{t-1} + \cdots + \beta_q \Delta m_{t-q} + \varepsilon_{1t}, \quad \textbf{(23)}$$

$$\Delta m_t = \alpha_0 + \alpha_1 \Delta m_{t-1} + \cdots + \alpha_p \Delta m_{t-p} + \beta_1 \Delta ipi_{t-1} + \cdots + \beta_q \Delta ipi_{t-q} + \varepsilon_{2t}, \quad \textbf{(24)}$$

where Δm denotes $\Delta m1$, $\Delta m2$, or Δir.

In comparison, $\widehat{SM}^{(1)}$ checks the hypotheses of no Granger causality in mean:

$$E\left(\Delta ipi_t | \Delta m_{t-q}^{t-1}, \Delta ipi_{t-p}^{t-1}\right) = E\left(\Delta ipi_t | \Delta ipi_{t-p}^{t-1}\right),$$

$$E\left(\Delta m_t | \Delta ipi_{t-q}^{t-1}, \Delta m_{t-p}^{t-1}\right) = E\left(\Delta m_t | \Delta m_{t-p}^{t-1}\right),$$

and $\widehat{SM}$ checks the hypotheses of no Granger causality in distribution:

$$f\left(\Delta ipi_t, \Delta m_{t-q}^{t-1} | \Delta ipi_{t-p}^{t-1}\right) = f\left(\Delta ipi_t | \Delta ipi_{t-p}^{t-1}\right) f\left(\Delta m_{t-q}^{t-1} | \Delta ipi_{t-p}^{t-1}\right),$$

$$f\left(\Delta m_t, \Delta ipi_{t-q}^{t-1} | \Delta m_{t-p}^{t-1}\right) = f\left(\Delta m_t | \Delta m_{t-p}^{t-1}\right) f\left(\Delta ipi_{t-q}^{t-1} | \Delta m_{t-p}^{t-1}\right),$$

where $\Delta m_{t-s}^{t-1} = (\Delta m_{t-1}, \ldots, \Delta m_{t-s})$, and $\Delta ipi_{t-s}^{t-1} = (\Delta ipi_{t-1}, \ldots, \Delta ipi_{t-s})$, with $s = p, q$. The $\widehat{SM}^{(1)}$ test checks whether past money (output) growths are useful in predicting the mean of future output (money) growths. The $\widehat{SM}$ test checks whether past money (output) growths are useful in predicting the distribution of future output (money) growths. We note that density forecasts for macroeconomic variables have been important for such decision makers as central banks (Diebold, Hahn, and Tay, 1999; Clements, 2004; Casillas-Olvera and Bessler, 2006).

We apply $\widehat{SM}$, Su and White's (2007) test, $\widehat{SM}^{(1)}$ and Granger's (1969) F test to investigate various Granger causalities between output and money. All the data are standardized to have zero mean and unit variance. For $\widehat{SM}$ and $\widehat{SM}^{(1)}$, we use the Gaussian kernel, the weighting function $a(X_t) = \mathbf{1}(|X_t| \leq 1.5)$, and the $N(0,1)$ CDF for $W_1(\cdot)$ and $W_2(\cdot)$. We set the bandwidth $h = h^* n^{-3/[2(4+d_x)]}$, where h^* is the least squares cross-validated (LSCV) bandwidth for estimating $E(Y_t|X_t)$. For Su and White's (2007) test, we use the same kernel function and bandwidths as Su and White (2007), i.e., the fourth order kernel, and $h_1 = \tilde{h}^* n^{1/(8+d_x+d_z)-1/(4+d_x+d_z)}$, $h_2 = h^* n^{1/(8+d_x)-1/(4+d_x)}$, where $\tilde{h}^*$ is the LSCV bandwidth for estimating $E(Y_t|X_t, Z_t)$. We choose the Gaussian kernel as the bootstrap kernel, and set the resampling bandwidth $h_b = n^{-1/[d_x(d_x+4)]}$, which satisfies Assumption $A.8$ in Paparoditis and Politis (2000). We use $B = 200$ bootstrap iterations and obtain the LSCV bandwidths in each iteration. We also consider the following two bandwidths: (1) fix $h^* = \tilde{h}^* = d_x$ for both the original data and bootstrap samples; (2) select the LSCV bandwidths h^* and $\tilde{h}^*$ using the original data and hold them fixed in bootstrap iterations. The results are similar to Table 3 and are not reported here.

TABLE 3. Granger causality tests between money and output

	$\mathbb{H}_0$: Δipi does not Granger cause Δm						$\mathbb{H}_0$: Δm does not Granger cause Δipi					
	Δm_1	Δm_2	Δir	Δm_1	Δm_2	Δir	Δm_1	Δm_2	Δir	Δm_1	Δm_2	Δir
	LIN			$\widehat{SM}^{(1)}$			*LIN*			$\widehat{SM}^{(1)}$		
$p=1, q=1$	0.798	0.557	0.000	0.165	0.220	0.055	0.332	0.491	0.034	0.430	0.625	0.035
$p=1, q=2$	0.144	0.232	0.000	0.000	0.175	0.010	0.581	0.739	0.052	0.685	0.295	0.015
$p=1, q=3$	0.000	0.098	0.000	0.000	0.025	0.020	0.684	0.329	0.111	0.385	0.145	0.085
$p=2, q=1$	0.848	0.466	0.000	0.085	0.175	0.005	0.364	0.432	0.094	0.660	0.660	0.145
$p=2, q=2$	0.282	0.232	0.000	0.000	0.085	0.000	0.662	0.679	0.188	0.830	0.520	0.130
$p=2, q=3$	0.000	0.089	0.000	0.000	0.035	0.000	0.722	0.341	0.296	0.400	0.340	0.150
$p=3, q=1$	0.855	0.413	0.000	0.105	0.180	0.010	0.280	0.336	0.139	0.500	0.365	0.085
$p=3, q=2$	0.218	0.185	0.000	0.015	0.020	0.005	0.551	0.579	0.302	0.690	0.230	0.060
$p=3, q=3$	0.000	0.071	0.000	0.000	0.025	0.010	0.714	0.286	0.349	0.285	0.160	0.085
	SW07			$\widehat{SM}$			*SW07*			$\widehat{SM}$		
$p=1, q=1$	0.345	0.190	0.000	0.025	0.075	0.015	0.855	0.525	0.000	0.425	0.130	0.000
$p=1, q=2$	0.570	0.825	0.015	0.000	0.075	0.020	0.890	0.265	0.710	0.450	0.015	0.000
$p=1, q=3$	0.185	0.975	0.055	0.000	0.135	0.040	0.960	0.840	0.825	0.150	0.025	0.010
$p=2, q=1$	0.370	0.230	0.005	0.000	0.045	0.000	0.070	0.050	0.005	0.415	0.145	0.000
$p=2, q=2$	0.110	0.540	0.000	0.000	0.040	0.000	0.645	0.715	0.815	0.310	0.045	0.000
$p=2, q=3$	0.315	0.870	0.015	0.000	0.025	0.000	0.530	0.875	0.700	0.150	0.045	0.005
$p=3, q=1$	0.000	0.065	0.005	0.050	0.050	0.000	0.790	0.290	0.080	0.520	0.110	0.000
$p=3, q=2$	0.095	0.170	0.025	0.010	0.005	0.000	0.115	0.275	0.165	0.595	0.020	0.000
$p=3, q=3$	0.135	0.405	0.015	0.005	0.015	0.000	0.025	0.260	0.545	0.105	0.010	0.000

Notes: (i) *LIN* and *SW07* denote Granger's (1969) F test and Su and White's (2007) test; (ii) numbers in the main entries are p-values; (iii) the p-values of Granger's (1969) test are calculated using the $F_{q,n-p-q}$ distribution; (iv) the p-values of Su and White's (2007) test and the $\widehat{SM}$, $\widehat{SM}^{(1)}$ tests are based on 200 bootstrap iterations.

The top half of Table 3 reports the results of Granger's (1969) F test and the $\widehat{SM}^{(1)}$ test. For Granger's (1969) F test, we observe that at the 5% level, all three monetary variables do not Granger-cause output, suggesting ineffectiveness of monetary policy. This is consistent with Uhlig's (2005) linear VAR based conclusion that monetary policy shocks have no clear effect on real GDP. The results of Granger's (1969) F test also suggest that the growth of M2 does not respond to the growth of output, and the growth of M1 responds to the growth of output only at the third order lag. However, it rejects the null hypothesis that Δipi does not Granger-cause Δir at any lag, which may indicate the existence of a linear Taylor rule (Taylor, 1993). Compared with Granger's (1969) F test, $\widehat{SM}^{(1)}$ reveals stronger evidence of Granger causalities in mean from output to money. Based on $\widehat{SM}^{(1)}$, Δipi Granger-causes $\Delta m1$ at the second and third order lags, and Granger-causes $\Delta m2$ at the third order lag. Thus, $\widehat{SM}^{(1)}$ documents the existence of nonlinear Granger causalities in mean and provides justification for modeling the relationship between money and output by a nonlinear conditional mean model.

The bottom half of Table 3 reports the results of Su and White's (2007) test and the $\widehat{SM}$ test. Comparing the results of Su and White's (2007) test with those of Granger's (1969) F test, we find no significant difference between them. That is, Su and White's (2007) test cannot detect any additional relationship between money and output beyond Granger's (1969) F test. However, $\widehat{SM}$ documents strong evidence against the hypothesis that output does not Granger-cause money for all three monetary variables and at all lag orders, except for $\Delta m2$ at $p = 1$. This implies that the Federal Reserve Board responds to economic conditions and uses monetary polices to stimulate recovery or curb overheating. Moreover, $\widehat{SM}$ shows that interest rate is effective in stimulating the economy at all lag orders, and there is one month lag for M2 to affect output. We find no evidence against ineffectiveness of M1 in affecting the economy, and it might not be difficult to understand this given the development of financial markets and direct financing. To sum up, the results of $\widehat{SM}$ indicate strong evidence against distributional non-Granger causality between money and output. This is consistent with the recent use of nonlinear models to capture the relationship between money and output in the literature.

We note that we find Granger causality in mean from output to broad money supply (with a p-value of 0.025), but do not find Granger causality in distribution for the case of $p = 1, q = 3$ (with a p-value of 0.135). This contradiction might be caused by the power loss of testing Granger causality in distribution when the DGP is a Granger causality in mean process. That is, if the true DGP is a Granger causality in mean process, then $\widehat{SM}^{(1)}$ will be more powerful than $\widehat{SM}$. This is because $\widehat{SM}$ checks the whole distribution, which could be viewed as checking a weighted average of Granger causality effects in all moments and so is not efficient when the truth is a Granger causality in mean process.

Moreover, the results of $\widehat{SM}^{(1)}$ and $\widehat{SM}$ reveal the existence of Granger causality in distribution rather than Granger causality in mean between money and output for most cases. The distributional effectiveness of monetary policy implies that the effects of monetary policy may be underestimated by a conditional mean model. Our results are consistent with Lee and Yang (2012), who document Granger causality in an output-money quantile relationship, particularly in the tails of the output growth distribution. This may indicate that the Federal Reserve Board intervenes in the economy when it is overheated or is in recession, and remains idle when the economy grows moderately. From the evidence it appears, the distributional Granger causality relationship provides a more comprehensive picture on the role of monetary policy.

9. CONCLUSION

Conditional independence is a widely maintained condition and encompasses many important hypotheses in econometrics and statistics, such as the Markov property, non-Granger causality, missing at random and exogeneity. In this paper, we have proposed a new test for conditional independence via a nonparametric regression approach in combination with the use of conditional characteristic function. Our nonparametric regression approach makes the test only involve smoothing over conditioning variables. As a result, the test alleviates the effect of the curse of dimensionality problem and is asymptotically more powerful than the existing smoothed nonparametric tests for conditional independence under a class of local alternatives. Moreover, by using a single bandwidth, we allow the nonparametric estimators of both conditional joint and marginal characteristic functions to jointly determine the asymptotic distribution of the test. Because of better asymptotic approximation, the test has significantly better size than the existing smoothed nonparametric tests in finite samples. On the other hand, the use of the conditional characteristic function allows us to infer patterns of conditional dependence. By taking appropriate partial derivatives, our approach can generate a variety of model-free tests for omitted variables, Granger causality in various moments, and conditional uncorrelatedness respectively. In particular, the derivative test for omitted variables is asymptotically more powerful than the smoothed nonparametric tests of Fan and Li (1996), Lavergne and Vuong (2000) and Aït-Sahalia et al. (2001).

Simulations show that in comparison with Granger's (1969) F test and Nishiyama et al.'s (2011) test for Granger causality in mean, and Su and White's (2007, 2012) tests for Granger causality in distribution, the proposed tests have reasonable size and excellent power against various alternatives in finite samples. We apply our tests to investigate Granger causality between money and output, and document strong evidence on some nonlinear relationships. Our findings justify the use of nonlinear models for linking money and output.

Our approach could be extended in several directions. For example, we might test conditional independence between Y and Z given X when Y is only partially observed, or to test the hypothesis of constant conditional dependence, i.e., the joint dependence of (Y, Z) given X does not depend on the values of X (noting that conditional independence almost everywhere does not imply independence (Phillips, 1988)). We may also use the characteristic function to test strict stationarity by replacing the conditioning random variable X with deterministic time t or normalized time t/n and considering whether a finite-dimensional distribution of a time series changes over time. All these problems will be pursued in subsequent studies.

Supplementary material to this article is provided in "Supplementary Material to 'Characteristic Function Based Testing for Conditional Independence: A Nonparametric Regression Approach'", which is available at Cambridge Journal Online (journals.cambridge.org/ect).

REFERENCES

Aït-Sahalia, Y., P.J. Bickel, & T.M. Stoker (2001) Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics* 105, 363–412.

Bae, Y. & R. de Jone (2007) Money demand function estimation by nonlinear cointegration. *Journal of Applied Econometrics* 22, 767–793.

Blundell, R. & J.L. Horowitz (2007) A non-parametric test of exogeneity. *Review of Economic Studies* 74, 1035–1058.

Bouezmarni, T., J.V.K. Rombouts, & A. Taamouti (2012) Nonparametric copula-based test for conditional independence with applications to granger causality. *Journal of Business and Economic Statistics* 30, 275–287.

Bouezmarni, T. & A. Taamouti (2014) Nonparametric tests for conditional independence using conditional distributions. *Journal of Nonparametric Statistics* 26, 697–719.

Campbell, J.Y. (1992) No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31, 281–318.

Casillas-Olvera, G. & D.A. Bessler (2006) Probability forecasting and central bank accountability. *Journal of Policy Modeling* 28, 223–234.

Chen, B. & Y. Hong (2010) Characteristic function-based testing for multifactor continuous-time markov models via nonparametric regression. *Econometric Theory* 26, 1115–1179.

Chen, R. & R.S. Tsay (1993) Functional-coefficient autoregressive models. *Journal of American Statistical Association* 88, 298–308.

Christiano, L.J. & L. Ljungqvist (1988) Money does Granger-cause output in the bivariate money-output relation. *Journal of Monetary Economics* 22, 217–235.

Clements, M.P. (2004) Evaluating the Bank of England density forecasts of inflation. *Economic Journal* 114, 844–866.

Dawid, A.P. (1979) Conditional independence in statistical theory. *Journal of Royal Statistical Society: Series B* 41, 1–31.

Dawid, A.P. (1980) Conditional independence for statistical operations. *Annals of Statistics* 8, 598–617.

Diebold, F.X., J. Hahn, & A.S. Tay (1999) Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns of foreign exchange. *Review of Economics and Statistics* 81, 661–673.

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

Engle, R.F., D.M. Lilien, & R.P. Robins (1987) Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica* 55, 391–407.

Fan, Y. & I. Gijbels (1996) *Local Polynomial Modelling and its Applications*. Chapman and Hall.
Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted varibales and semiparametric function forms. *Econometrica* 64, 865–890.
Fan, Y. & Q. Li (1999) Root$-n-$consistent estimation of partially linear time series models. *Journal of Nonparametric Statistics* 10, 245–271.
Friedman, B.M. & K.N. Kutter (1993) Another look at the evidence on money-income causality. *Journal of Econometrics* 57, 189–203.
Gao, J. & I. Gijbels (2008) Bandwidth selection in nonparametric kernel testing. *Journal of American Statistical Association* 103, 1584–1594.
Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.
Granger, C.W.J. (1980) Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control* 2, 329–352.
Hahn, J., P. Todd, & W.V. Klaauw (2001) Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69, 201–209.
Heckman, J.D. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5, 475–492.
Hjellvik, V., Q. Yao, & D. Tjøstheim (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.
Hong, Y. (1999) Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of American Statistical Association* 94, 1201–1220.
Horowitz, J.L. & C.F. Manski (2000) Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of American Statistical Association* 95, 77–84.
Horowitz, J.L. & V.G. Spokoiny (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–631.
Huang, T.M. (2010) Testing conditional independence using maximal nonlinear conditional correlation. *Annals of Statistics* 38, 2047–2091.
Huang, M., Y. Sun, & H. White (2016) A flexible nonparametric test for conditional independence, *Econometric Theory* 32, 1–49.
Kandil, M. (1995) Asymmetric norminal flexibility and economic fluctuations. *Southern Economic Journal* 61, 674–695.
Kim, C.J. & C.R. Nelson (2006) Estimation of a forward-looking monetary policy rule: A time-varying parameter model using ex post data. *Journal of Monetary Economics* 53, 1949–1966.
Lavergne, P. & Q. Vuong (2000) Nonparametric siginificance testing. *Econometric Theory* 16, 576–601.
Lee, J. (2013) A Consistent Nonparametric Bootstrap Test of Exogeneity. Working paper, University of St Andrews.
Lee, T. & W. Yang (2012) Money-income granger-causality in quantiles. In D. Millimet & D. Terrell (eds.), *Advances in Econometrics*, vol. 30, pp. 383–407. Emerald Publishers.
Linton, O. & P. Gozalo (2014) Testing conditional independence restrictions. *Econometric Reviews* 33, 523–552.
Little, R.J.A. (1985) A note about models for selectivity bias. *Econometrica* 53, 1469–1474.
Manski, C.F. (2000) Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice. *Journal of Econometrics* 95, 415–442.
Manski, C.F. (2003) *Partial Identification of Probability Distribution*. Springer-Verlag.
Manski, C.F. (2007) *Identification for Prediction and Decision*. Princeton University Press.
Masry, E. (1996a) Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571–599.

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

Nishiyama, Y., K. Hitomi, Y. Kawasaki, & K. Jeong (2011) A consistent nonparametric test for nonlinear causalityłspecification in time series regression. *Journal of Econometrics* 165, 112–127.

Nobay, R.A. & D.A. Peel (2003) Optimal discretionary monetary policy in a model of asymmetric central bank preferences. *Economic Journal* 113, 657–665.

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

Peiró, A. (1999) Skewness in financial returns. *Journal of Banking and Finance* 23, 847–862.

Phillips, P.C.B. (1988) Conditional and unconditional statistical independence. *Journal of Econometrics* 38, 341–348.

Priestley, M.B. (1988) *Nonlinear and Nonstationary Time Series Analysis*. Academic Press.

Psaradakis, Z., M.O. Ravn, & M. Sola (2005) Markov switching causality and the money-output relationship. *Journal of Applied Econometrics* 20, 665–683.

Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Annals of Statistics* 3, 1–14.

Rubin, D.B. (1976) Inference and missing data. *Biometrika* 63, 581–592.

Rust, J. (1994) Structural estimation of markov decision processes. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 3081–3143. Elsevier.

Sims, C.A. (1972) Money, income, and causality. *American Economic Review* 62, 540–552.

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

Song, K. (2009) Testing conditional independence via rosenblatt transforms. *Annals of Statistics* 37, 4011–4045.

Stock, J.H. & M.W. Watson (1989) Interpretating the evidence on money-income causality. *Journal of Econometrics* 40, 161–181.

Su, L. & A. Ullah (2009) Testing conditional uncorrelatedness. *Journal of Business and Economic Statistics* 27, 18–29.

Su, L. & H. White (2007) A consistent characteristic function-based test for conditional independence. *Journal of Econometrics* 141, 807–834.

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

Su, L. & H. White (2012) Conditional independence specification testing for dependent process with local polynomial quantile regression. *Advances in Econometrics* 29, 355–434.

Su, L. & H. White (2014) Testing conditional independence via empirical likelihood. *Journal of Econometrics* 182, 27–44.

Sun, Y., P. Phillips, & S. Jin (2008) Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* 76, 175–194.

Taamouti, A., T. Bouezmarni, & A. El Ghouch (2014) Nonparametric estimation and inference for conditional density based granger causality measures. *Journal of Econometrics* 180, 251–264.

Taylor, J. (1993) Discretion versus policy rules in practice. *Carnegie-Rochester conference Series on Public Policy* 39, 195–214.

Tenreiro, C. (1997) Loi asymptotique des erreurs quadratiques intégrées des estimateurs a noyau de la densité et de la régression sou des conditions de dependance. *Portugaliae Mathematica* 54, 187–213.

Tong, H. & K.S. Lim (1980) Threshold autoregression, limit cycles and cyclical data. *Journal of Royal Statistical Society, Series B* 42, 245–292.

Uhlig, H. (2005) What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics* 52, 381–419.

Wang, Q., O. Linton, & W. Hardle (2004) Semiparametric regression analysis with missing response at random. *Journal of American Statistical Association* 99, 334–345.

MATHEMATICAL APPENDIX

Throughout the appendix, we denote

$$\hat{M}_h = nh^{d_x/2}\hat{M} = h^{d_x/2}\sum_{t=1}^{n}\iint |\hat{\sigma}(u,v,X_t)|^2 a(X_t)dW_1(u)dW_2(v),$$

and $\varepsilon_{yz}(u,v,X_s) = e^{\mathbf{i}(u'Y_s+v'Z_s)} - \phi_{yz}(u,v,X_s), \varepsilon_y(u,X_s) = \varepsilon_{yz}(u,0,X_s), \varepsilon_z(v,X_s) = \varepsilon_{yz}(0,v,X_s)$. In addition, $\xi_t = (X_t', Y_t', Z_t')'$, $C \in (0,\infty)$ is a generic bounded constant that may vary from case to case, A^* denotes the conjugate of A, and $Re(A)$ denotes the real part of A.

Proof of Theorem 1. Under $\mathbb{H}_0 : \phi_{yz}(u,v,x) = \phi_y(u,x)\phi_z(v,x)$, we can decompose $\hat{\sigma}(u,v,x)$ as follows:

$$\hat{\sigma}(u,v,x) = \left[\hat{\phi}_{yz}(u,v,x) - \phi_{yz}(u,v,x)\right] - \phi_z(v,x)\left[\hat{\phi}_y(u,x) - \phi_y(u,x)\right] - \phi_y(u,x)\left[\hat{\phi}_z(v,x) - \phi_z(v,x)\right] - \left[\hat{\phi}_y(u,x) - \phi_y(u,x)\right]\left[\hat{\phi}_z(v,x) - \phi_z(v,x)\right]. \quad \textbf{(A.1)}$$

According to $(A.1)$, we decompose $\hat{M}_h$ as follows:

$$\begin{aligned}\hat{M}_h = h^{d_x/2}\sum_{t=1}^{n}\iint &\Big\{\left|\hat{\phi}_{yz} - \phi_{yz}\right|^2 + \left|\phi_y\right|^2\left|\hat{\phi}_z - \phi_z\right|^2 + \left|\phi_z\right|^2\left|\hat{\phi}_y - \phi_y\right|^2 \\ &+ 2Re\left[\phi_y\phi_z^*(\hat{\phi}_z - \phi_z)(\hat{\phi}_y - \phi_y)^*\right] - 2Re\left[(\hat{\phi}_{yz} - \phi_{yz})\phi_y^*(\hat{\phi}_z - \phi_z)^*\right] \\ &- 2Re\left[(\hat{\phi}_{yz} - \phi_{yz})\phi_z^*(\hat{\phi}_y - \phi_y)^*\right] + \left|(\hat{\phi}_y - \phi_y)(\hat{\phi}_z - \phi_z)\right|^2 \\ &- 2Re\left[(\hat{\phi}_{yz} - \phi_{yz})(\hat{\phi}_y - \phi_y)^*(\hat{\phi}_z - \phi_z)^*\right] + 2Re\left[(\hat{\phi}_y - \phi_y)\phi_y^*\right]\left|\hat{\phi}_z - \phi_z\right|^2 \\ &+ 2Re\left[(\hat{\phi}_z - \phi_z)\phi_z^*\right]\left|\hat{\phi}_y - \phi_y\right|^2\Big\} a(X_t)dW_1(u)dW_2(v) = \sum_{i=1}^{10} T_i, \text{ say,} \end{aligned} \quad \textbf{(A.2)}$$

where $\hat{\phi}_{yz} \equiv \hat{\phi}_{yz}(u,v,X_t)$, $\hat{\phi}_y \equiv \hat{\phi}_y(u,X_t)$, $\hat{\phi}_z \equiv \hat{\phi}_z(v,X_t)$, $\phi_{yz} \equiv \phi_{yz}(u,v,X_t)$, $\phi_y \equiv \phi_y(u,X_t)$, $\phi_z \equiv \phi_z(v,X_t)$. We shall analyze each of these terms $\{T_i\}_{i=1}^{10}$ in $(A.2)$ to identify the leading terms that determine the asymptotic distribution of our test statistic. The leading terms are given by Propositions A.1–A.7 below.

PROPOSITION A.1. *Under the conditions of Theorem 1,* $T_1 = B_1 + \tilde{U}_1 + o_P(1)$, *where*

$$B_1 = h^{-d_x/2}\iiint\left[1 - \left|\phi_{yz}(u,v,x)\right|^2\right]dW_1(u)dW_2(v)a(x)dx\int D^2(p,\tau)K(\tau)^2d\tau,$$

$$\tilde{U}_1 = \frac{2}{nh^{3d_x/2}}\sum_{1\le s<r\le n} U_1(\xi_s,\xi_r),$$

and $U_1(\xi_s,\xi_r) = \iiint \frac{a(x)}{g(x)}D\left(p,\frac{X_s-x}{h}\right)D\left(p,\frac{X_r-x}{h}\right)K\left(\frac{X_s-x}{h}\right)K\left(\frac{X_r-x}{h}\right) Re\left[\varepsilon_{yz}(u,v,X_s)\varepsilon_{yz}(u,v,X_r)^*\right]dW_1(u)dW_2(v)dx.$

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

PROPOSITION A.2. *Under the conditions of Theorem 1,* $T_2 = B_2 + \tilde{U}_2 + o_P(1)$, *where*

$$B_2 = h^{-d_x/2} \iiint |\phi_y(u,x)|^2 \left[1 - |\phi_z(v,x)|^2\right] dW_1(u) dW_2(v) a(x) dx \int D^2(p,\tau) K(\tau)^2 d\tau,$$

$$\tilde{U}_2 = \frac{2}{nh^{3d_x/2}} \sum_{1 \le s < r \le n} U_2(\xi_s, \xi_r),$$

and $U_2(\xi_s,\xi_r) = \iiint \frac{a(x)}{g(x)} D\left(p, \frac{X_s - x}{h}\right) D\left(p, \frac{X_r - x}{h}\right) K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right) |\phi_y(u,x)|^2$ $Re\left[\varepsilon_z(v, X_s)\varepsilon_z(v, X_r)^*\right] \times dW_1(u) dW_2(v) dx.$

PROPOSITION A.3. *Under the conditions of Theorem 1,* $T_3 = B_3 + \tilde{U}_3 + o_P(1)$, *where*

$$B_3 = h^{-d_x/2} \iiint |\phi_z(v,x)|^2 \left[1 - \left|\phi_y(u,x)\right|^2\right] dW_1(u) dW_2(v) a(x) dx \int D^2(p,\tau) K(\tau)^2 d\tau,$$

$$\tilde{U}_3 = \frac{2}{nh^{3d_x/2}} \sum_{1 \le s < r \le n} U_3(\xi_s, \xi_r),$$

and $U_3(\xi_s,\xi_r) = \iiint \frac{a(x)}{g(x)} D\left(p, \frac{X_s - x}{h}\right) D\left(p, \frac{X_r - x}{h}\right) K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right) |\phi_z(v,x)|^2$ $Re\left[\varepsilon_y(u, X_s)\varepsilon_y(u, X_r)^*\right] \times dW_1(u) dW_2(v) dx.$

PROPOSITION A.4. *Under the conditions of Theorem 1,* $T_4 = \tilde{U}_4 + o_P(1)$, *where* $\tilde{U}_4 = \frac{2}{nh^{3d_x/2}} \sum_{s \ne r} U_4(\xi_s, \xi_r)$, *and* $U_4(\xi_s,\xi_r) = \iiint \frac{a(x)}{g(x)} D\left(p, \frac{X_s - x}{h}\right) D\left(p, \frac{X_r - x}{h}\right)$ $K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right) Re\left[\phi_y(u,x)\phi_z(v,x)^* \varepsilon_z(v, X_s)\varepsilon_y(u, X_r)^*\right] dW_1(u) dW_2(v) dx.$

PROPOSITION A.5. *Under the conditions of Theorem 1,* $T_5 = B_5 + \tilde{U}_5 + o_P(1)$, *where*

$$B_5 = -2h^{-d_x/2} \iiint |\phi_y(u,x)|^2 \left[1 - |\phi_z(v,x)|^2\right] dW_1(u) dW_2(v) a(x) dx \int D^2(p,\tau) K(\tau)^2 d\tau,$$

$$\tilde{U}_5 = \frac{2}{nh^{3d_x/2}} \sum_{s \ne r} U_5(\xi_s, \xi_r),$$

and $U_5(\xi_s,\xi_r) = -\iiint \frac{a(x)}{g(x)} D\left(p, \frac{X_s - x}{h}\right) D\left(p, \frac{X_r - x}{h}\right) K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right)$ $Re\left[\phi_y(u,x)^* \varepsilon_{yz}(u,v,X_s)\varepsilon_z(v, X_r)^*\right] dW_1(u) dW_2(v) dx.$

PROPOSITION A.6. *Under the conditions of Theorem 1,* $T_6 = B_6 + \tilde{U}_6 + o_P(1)$, *where*

$$B_6 = -2h^{-d_x/2} \iiint |\phi_z(v,x)|^2 \left[1 - \left|\phi_y(u,x)\right|^2\right] dW_1(u) dW_2(v) a(x) dx \int D^2(p,\tau) K(\tau)^2 d\tau,$$

$$\tilde{U}_6 = \frac{2}{nh^{3d_x/2}} \sum_{s \ne r} U_6(\xi_s, \xi_r),$$

and $U_6(\xi_s,\xi_r) = -\iiint \frac{a(x)}{g(x)} D\left(p, \frac{X_s - x}{h}\right) D\left(p, \frac{X_r - x}{h}\right) K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right)$ $Re\left[\phi_z(v,x)^* \varepsilon_{yz}(u,v,X_s)\varepsilon_y(u, X_r)^*\right] dW_1(u) dW_2(v) dx.$

PROPOSITION A.7. *Under the conditions of Theorem 1,* $T_7 + T_8 + T_9 + T_{10} = o_P(1)$.

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

Based on Propositions A.1–A.7, we can obtain the asymptotic centering factor B, and the leading term U that determines the asymptotic distribution of the test statistic:

$$B = \sum_{i=1}^{6} B_i = h^{-d_x/2} \iiint \left[1 - |\phi_y(u,x)|^2\right]\left[1 - |\phi_z(v,x)|^2\right] dW_1(u) dW_2(v) a(x) dx \times \int K^2(\tau) D^2(\tau) d\tau, \quad \textbf{(A.3)}$$

$$U = \sum_{i=1}^{6} \tilde{U}_i = \frac{2}{nh^{3d_x/2}} \sum_{1 \le s < r \le n} U(\xi_s, \xi_r),$$

where $U(\xi_s,\xi_r) = \sum_{i=1}^{3} U_i(\xi_s,\xi_r) + 2\sum_{i=4}^{6} U_i(\xi_s,\xi_r)$.

Proposition A.8 provides the asymptotic distribution of U, which is a second order degenerate U-statistic.

PROPOSITION A.8. *Under the conditions of Theorem 1,* $U/\sqrt{V} \xrightarrow{d} N(0,1)$*, where*

$$V = 2\int \left[\iint |\Phi_y(u_1+u_2,x)|^2 dW_1(u_1) dW_1(u_2) \iint |\Phi_z(v_1+v_2,x)|^2 dW_2(v_1) dW_2(v_2) \right] a^2(x) dx \times \int \left[\int D(p,\tau) D(p,\tau+\eta) K(\tau) K(\tau+\eta) d\tau \right]^2 d\eta, \quad \textbf{(A.4)}$$

with $\Phi_s(a_1+a_2,x) = \phi_s(a_1+a_2,x) - \phi_s(a_1,x)\phi_s(a_2,x)$ *for* $s = y$ *or* z.

As the test is obtained by replacing the asymptotic centering factor B and the scaling factor V by their estimators $\hat{B}$ and $\hat{V}$, which are given in (13) and (14) respectively, we shall show that replacing B and V by $\hat{B}$ and $\hat{V}$ does not affect the limiting distribution of the test statistic.

PROPOSITION A.9. *Under the conditions of Theorem 1,* $\hat{B} - B = o_P(1)$ *and* $\hat{V} - V = o_P(1)$.

The proof of Theorem 1 will be completed provided Propositions A.1–A.9 are proven, which we turn to next. Since the proofs of Propositions A.1–A.7 are rather similar, for space we only focus on the proofs of Propositions A.8 and A.9. A detailed proof for all Propositions 1–9 is provided in the online Supplementary Material. ■

Proof of Proposition A.8. Because $E[U(\xi_s,\xi)] = E[U(\xi',\xi_r)] = 0$ for any given ξ and ξ', $U = \sum_{1\le s<r\le n} U(\xi_s,\xi_r)$ is a second order degenerate U-statistic. Following Tenreiro's (1997) central limit theorem (the English version of this theorem has been stated by Su and White, 2008, Theorem A.4, pp. 852–853.) for degenerate U-statistics in a time series context, we have $\sigma_n^{-1} \sum_{1\le s<r\le n} U(\xi_s,\xi_r) \xrightarrow{d} N(0,1)$ if the following conditions hold: For some constants $\delta_0 > 0$, $\gamma_0 < \frac{1}{2}$ and $\gamma_1 > 0$, (i) $u_n(4+\delta_0) = O(n^{\gamma_0})$; (ii) $v_n(2) = o(1)$; (iii) $w_n(2+\delta_0/2) = o(n^{1/2})$, and (iv) $z_n(2)n^{\gamma_1} = O(1)$, where $\sigma_n^2 = \sum_{1\le s<r\le n} \mathrm{var}[U(\xi_s,\xi_r)]$,

$$u_n(p) = \max\left\{ \max_{1\le i\le n} \|U(\xi_i,\xi_1)\|_p, \|U(\xi_1,\bar{\xi}_1)\|_p \right\}, \quad v_n(p) = \max\left\{ \max_{1\le i\le n} \|G_{n1}(\xi_i,\xi_1)\|_p, \|G_{n1}(\xi_1,\bar{\xi}_1)\|_p \right\},$$

$$w_n(p) = \|G_{n1}(\xi_1,\xi_1)\|_p, \quad z_n(p) = \max_{1\le i\le n} \max_{1\le j\le n} \left\{ \|G_{nj}(\xi_i,\xi_1)\|_p, \|G_{nj}(\xi_1,\xi_i)\|_p, \|G_{nj}(\xi_1,\bar{\xi}_1)\|_p \right\},$$

$G_{ni}(\eta,\tau) = E\left[U(\xi_i,\eta)U(\xi_1,\tau)\right]$, $\bar{\xi}_1$ is an independent copy of ξ_1, and $\|\cdot\|_p = \left\{E|\cdot|^p\right\}^{1/p}$ for $p \geq 1$.

First, we calculate the asymptotic variance of $U(\xi_s,\xi_r)$, namely $\sigma_0^2 = \text{var}[U(\xi_s,\xi_r)] = \iint U(\xi_s,\xi_r)^2 dP(\xi_s)dP(\xi_r)$. Since $U(\xi_s,\xi_r)$ contains six terms, we need to calculate the individual variances of these six terms as well as their fifteen pairwise covariances. In calculation, we have used the following facts: (1) under $\mathbb{H}_0$, Y_t is independent of Z_t conditional on X_t; (2) the weighting functions $W_1(u)$, $W_2(v)$ weigh sets symmetric about the origin equally, implying $\iint \Phi_y(u_1+u_2,x)dW_1(u_1)dW_1(u_2) = \iint \Phi_y(u_1-u_2,x) dW_1(u_1)dW_1(u_2)$, $\iint \Phi_z(v_1+v_2,x)dW_2(v_1)dW_2(v_2) = \iint \Phi_z(v_1-v_2,x) dW_2(v_1)dW_2(v_2)$; (3) similarly, $\iiiint \Phi_{yz}(u_1+u_2,v_1+v_2,x)dW_1(u_1) dW_1(u_2)dW_2(v_1)dW_2(v_2) = \iiiint \left[\phi_{yz}(u_1+u_2,v_1+v_2,x) - \phi_{yz}(u_1,v_1,x)\phi_{yz}(u_2,v_2,x)\right]dW_1(u_1)dW_1(u_2)dW_2(v_1)dW_2(v_2) = \iiiint \left[\phi_z(v_1+v_2,x)\Phi_y(u_1+u_2,x)+\phi_y(u_1+u_2,x)\Phi_z(v_1+v_2,x)\right]dW_1(u_1)dW_1(u_2)dW_2(v_1)dW_2(v_2)$. By tedious but straightforward algebra, we obtain

$$\sigma_0^2 = h^{3d_x}\int\left[\iint|\Phi_y(u_1+u_2,x)|^2 dW_1(u_1)dW_1(u_2)\iint|\Phi_z(v_1+v_2,x)|^2 dW_2(v_1)dW_2(v_2)\right]a^2(x)dx \times\int\left[\int D(p,\tau)D(p,\tau+\eta)K(\tau)K(\tau+\eta)d\tau\right]^2 d\eta.$$

It follows that $\sigma_n^2 = \sum_{1\leq s<r\leq n}\text{var}[U(\xi_s,\xi_r)] = \frac{n^2}{2}\sigma_0^2[1+o(1)]$. Hence, we have $V = \text{var}(U) = \frac{4}{n^2h^{3d_x}}\sigma_n^2 = \frac{2}{h^{3d_x}}\sigma_0^2[1+o(1)]$.

Now, we verify Conditions (i)–(iv). Since $U(\xi_s,\xi_r)$ is a sum of six terms, the product $U(\xi_i,\eta)U(\xi_j,\tau)$ contains 36 terms. All these terms have the same order of magnitude, and here we verify the first term $U_1(\xi_i,\eta)U_1(\xi_j,\tau)$ only. For $i \neq 1$,

$$\begin{aligned} E|U(\xi_i,\xi_1)|^q &\sim E\left|\iiint \frac{a(x)}{g(x)}D\left(p,\frac{X_i-x}{h}\right)D\left(p,\frac{X_1-x}{h}\right)K\left(\frac{X_i-x}{h}\right)K\left(\frac{X_1-x}{h}\right)\right.\\ &\quad\left.\times Re\left[\varepsilon_{yz}(u,v,X_i)\varepsilon_{yz}(u,v,X_1)^*\right]dW_1(u)dW_2(v)dx\right|^q\\ &= h^{qd_x}E\left|\iiint\frac{a(X_i-\tau h)}{g(X_i-\tau h)}D(p,\tau)D\left(p,\frac{X_1-X_i}{h}\right)K(\tau)K\left(\tau+\frac{X_1-X_i}{h}\right)\right.\\ &\quad\left.\times Re\left[\varepsilon_{yz}(u,v,X_i)\varepsilon_{yz}(u,v,X_1)^*\right]dW_1(u)dW_2(v)d\tau\right|^q = O\left(h^{(q+1)d_x}\right), \end{aligned}$$

so we have $\|U(\xi_i,\xi_1)\|_q = O\left(h^{d_x+d_x/q}\right)$. By a similar argument, we can obtain the same order of magnitude for $\|U(\xi_1,\bar{\xi}_1)\|_q$, where $\bar{\xi}_1$ is an independent copy of ξ_1. Hence, Condition (i) holds for any $\delta_0 > 0$ and $\gamma_0 < \frac{1}{2}$.

Next, we verify Condition (ii). Since for $i \neq 1$,

$$\begin{aligned} &E|G_{n1}(\xi_i,\xi_1)|^q = E|U(\xi_1,\xi_i)U(\xi_1,\xi_1)|^q\\ &\sim E\left|\iiint\frac{a(x)}{g(x)}D\left(p,\frac{X_1-x}{h}\right)D\left(p,\frac{X_i-x}{h}\right)K\left(\frac{X_1-x}{h}\right)K\left(\frac{X_i-x}{h}\right)\right.\\ &\times Re\left[\varepsilon_{yz}(u_1,v_1,X_1)\varepsilon_{yz}(u_1,v_1,X_i)^*\right]\\ &\times dW_1(u_1)dW_2(v_1)dx\iiint\frac{a(x')}{g(x')}D^2\left(p,\frac{X_1-x'}{h}\right)K^2\left(\frac{X_1-x'}{h}\right)\\ &\times\left.\left|\varepsilon_{yz}(u_2,v_2,X_1)\right|^2 dW_1(u_2)dW_2(v_2)dx'\right|^q \end{aligned}$$

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

$$= h^{2qd_x} E \left| \iiint \frac{a(X_1 - \tau h)}{g(X_1 - \tau h)} D(p,\tau) D\left(p, \tau + \frac{X_i - X_1}{h}\right) K(\tau) K\left(\tau + \frac{X_i - X_1}{h}\right) \right.$$
$$\times Re\left[\varepsilon_{yz}(u_1, v_1, X_1)\varepsilon_{yz}(u_1, v_1, X_i)^*\right]$$
$$\times d\tau dW_1(u_1) dW_2(v_1) \iiint \frac{a(X_1 - \eta h)}{g(X_1 - \eta h)} D^2(p,\eta) K^2(\eta)$$
$$\left. \times \left|\varepsilon_{yz}(u_2, v_2, X_1)\right|^2 dW_1(u_2) dW_2(v_2) d\eta \right|^q$$
$$= O\left(h^{(2q+1)d_x}\right),$$

we have $\|G_{n1}(\xi_i, \xi_1)\|_q = O\left(h^{(2+1/q)d_x}\right)$. By a similar argument, we can obtain the same order of magnitude for $\|G_{n1}(\xi_1, \bar{\xi}_1)\|_q$. Consequently, Condition (ii) is satisfied.

Now, we verify Condition (iii). Since

$$E\left|G_{n1}(\xi_1, \xi_1)\right|^q$$
$$= E\left|U(\xi_1, \xi_1) U(\xi_1, \xi_1)\right|^q$$
$$\sim E\left| \iiint \frac{a(x)}{g(x)} D^2\left(p, \frac{X_1 - x}{h}\right) K^2\left(\frac{X_1 - x}{h}\right) \left|\varepsilon_{yz}(u_1, v_1, X_1)\right|^2 dW_1(u_1) dW_2(v_1) dx \right.$$
$$\left. \times \iiint \frac{a(x')}{g(x')} D^2\left(p, \frac{X_1 - x'}{h}\right) K^2\left(\frac{X_1 - x'}{h}\right) \left|\varepsilon_{yz}(u_2, v_2, X_1)\right|^2 dW_1(u_2) dW_2(v_2) dx' \right|^q$$
$$= h^{2qd_x} E\left| \iiint \frac{a(X_1 - \tau h)}{g(X_1 - \tau h)} D^2(p,\tau) K^2(\tau) \left|\varepsilon_{yz}(u_1, v_1, X_1)\right|^2 dW_1(u_1) dW_2(v_1) d\tau \right.$$
$$\left. \times \iiint \frac{a(X_1 - \eta h)}{g(X_1 - \eta h)} D^2(p,\eta) K^2(\eta) \left|\varepsilon_{yz}(u_2, v_2, X_1)\right|^2 dW_1(u_2) dW_2(v_2) d\eta \right|^q$$
$$= O\left(h^{2qd_x}\right),$$

we have $w_n(q) = O\left(h^{2d_x}\right) = o(1)$. Hence, Condition (iii) holds.

To verify Condition (iv), for $i \neq j \neq 1$, we first calculate

$$E|G_{nj}(\xi_i, \xi_1)|^q$$
$$= E\left|E_j\left[U(\xi_j, \xi_i) U(\xi_1, \xi_1)\right]\right|^q$$
$$\sim E\left| \int \left[\iiint \frac{a(x)}{g(x)} D\left(p, \frac{X_j - x}{h}\right) D\left(p, \frac{X_i - x}{h}\right) K\left(\frac{X_j - x}{h}\right) K\left(\frac{X_i - x}{h}\right) \right.\right.$$
$$\times Re\left[\varepsilon_{yz}(u_1, v_1, X_j)\varepsilon_{yz}(u_1, v_1, X_i)^*\right]$$
$$\times dW_1(u_1) dW_2(v_1) dx \iiint \frac{a(x')}{g(x')} D^2\left(p, \frac{X_1 - x'}{h}\right) K^2\left(\frac{X_1 - x'}{h}\right)$$
$$\left.\left. \times \left|\varepsilon_{yz}(u_2, v_2, X_1)\right|^2 dW_1(u_2) dW_2(v_2) dx' \right] dP(\xi_j) \right|^q$$
$$= h^{2qd_x} E\left| \int \left[\iiint \frac{a(X_j - \tau h)}{g(X_j - \tau h)} D(p,\tau) D\left(p, \tau + \frac{X_i - X_j}{h}\right) K(\tau) K\left(\tau + \frac{X_i - X_j}{h}\right) \right.\right.$$
$$\times Re\left[\varepsilon_{yz}(u_1, v_1, X_j)\varepsilon_{yz}(u_1, v_1, X_i)^*\right] dW_1(u_1) dW_2(v_1) d\tau$$
$$\left.\left. \times \iiint \frac{a(X_1 - \eta h)}{g(X_1 - \eta h)} D^2(p,\eta) K^2(\eta) \left|\varepsilon_{yz}(u_2, v_2, X_1)\right|^2 dW_1(u_2) dW_2(v_2) d\eta \right] dP(\xi_j) \right|^q$$
$$= O\left(h^{3qd_x}\right).$$

By a similar argument, we can obtain $E|G_{nj}(\xi_1, \xi_i)|^q = O\left(h^{3qd_x + d_x}\right)$, $E|G_{nj}(\xi_1, \bar{\xi}_1)|^q = O\left(h^{3qd_x + d_x}\right)$. It follows that $z_n(p) = O\left(h^{3d_x}\right)$ and Condition (iv) holds by setting $\gamma_1 = 3\lambda d_x > 0$. The desired asymptotic normality follows immediately. ■

Proof of Proposition A.9. We should prove: (i) $\hat{B}-B=o_P(1)$, and (ii) $\hat{V}-V=o_P(1)$. Since the proofs of (i) and (ii) are similar, we focus on the proof of (i) here. We first decompose

$$
\begin{aligned}
&\hat{B}-B\\
&=h^{-d_x/2}\iiint a(x)\{[1-|\hat{\phi}_y(u,x)|^2][1-|\hat{\phi}_z(v,x)|^2]-[1-|\phi_y(u,x)|^2][1-|\phi_z(v,x)|^2]\}\\
&\quad\times dW_1(u)dW_2(v)dx\int D^2(p,\tau)K^2(\tau)d\tau\\
&=h^{-d_x/2}\iiint[|\phi_y(u,x)|^2-1][|\hat{\phi}_z(v,x)|^2-|\phi_z(v,x)|^2]dW_1(u)dW_2(v)a(x)dx\\
&\quad\times\int D^2(p,\tau)K^2(\tau)d\tau\\
&\quad+h^{-d_x/2}\iiint[|\phi_z(v,x)|^2-1][|\hat{\phi}_y(u,x)|^2-|\phi_y(u,x)|^2]dW_1(u)dW_2(v)a(x)dx\\
&\quad\times\int D^2(p,\tau)K^2(\tau)d\tau\\
&\quad+h^{-d_x/2}\iiint[|\hat{\phi}_y(u,x)|^2-|\phi_y(u,x)|^2][|\hat{\phi}_z(v,x)|^2-|\phi_z(v,x)|^2]dW_1(u)dW_2(v)a(x)dx\\
&\quad\times\int D^2(p,\tau)K^2(\tau)d\tau\\
&=H_1+H_2+H_3,\ \text{say.}
\end{aligned}
$$

To show $\hat{B}-B=o_P(1)$, we should prove $H_i=o_P(1)$ for $i=1,2,3$. Since the proofs of H_i, $i=1,2,3$, are rather similar, we focus on the proof of $H_1=o_P(1)$. We decompose H_1 as follows:

$$
\begin{aligned}
H_1&=h^{-d_x/2}\iiint[|\phi_y(u,x)|^2-1]|\hat{\phi}_z(v,x)-\phi_z(v,x)|^2dW_1(u)dW_2(v)a(x)dx\\
&\quad\times\int D^2(p,\tau)K^2(\tau)d\tau\\
&\quad+2h^{-d_x/2}\iiint[|\phi_y(u,x)|^2-1]Re\{[\hat{\phi}_z(v,x)-\phi_z(v,x)]\phi_z(v,x)^*\}dW_1(u)dW_2(v)a(x)dx\\
&\quad\times\int D^2(p,\tau)K^2(\tau)d\tau\\
&=H_{11}+H_{12},\ \text{say.}
\end{aligned}
$$

We further decompose H_{11} as follows:

$$
\begin{aligned}
H_{11}&\le 2h^{-d_x/2}\iiint[|\phi_y(u,x)|^2-1]\left|\hat{\phi}_z(v,x)-\bar{\phi}_z(v,x)\right|^2dW_1(u)dW_2(v)a(x)dx\\
&\quad\times\int D^2(p,\tau)K^2(\tau)d\tau\\
&\quad+2h^{-d_x/2}\iiint[|\phi_y(u,x)|^2-1]\left|\bar{\phi}_z(v,x)-\phi_z(v,x)\right|^2dW_1(u)dW_2(v)a(x)dx\\
&\quad\times\int D^2(p,\tau)K^2(\tau)d\tau\\
&=H_{11}^{(1)}+H_{11}^{(2)},\ \text{say,}
\end{aligned}
$$

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press

and

$$
\begin{aligned}
H_{11}^{(1)} &= \frac{2}{n^2 h^{5d_x/2}} \sum_{s=1}^{n} \iiint \left[|\phi_y(u,x)|^2 - 1\right] |\varepsilon_z(v, X_s)|^2 \, dW_1(u) dW_2(v) \\
&\quad \times \frac{a(x)}{g^2(x)} D^2\left(p, \frac{X_s - x}{h}\right) K^2\left(\frac{X_s - x}{h}\right) dx \int D^2(p,\tau) K^2(\tau) d\tau \\
&\quad + \frac{4}{n^2 h^{5d_x/2}} \sum_{l<s} \iiint \left[|\phi_y(u,x)|^2 - 1\right] Re\left[\varepsilon_z(v, X_s)\varepsilon_z(v, X_l)^*\right] dW_1(u) dW_2(v) \\
&\quad \times \frac{a(x)}{g^2(x)} D\left(p, \frac{X_s - x}{h}\right) D\left(p, \frac{X_l - x}{h}\right) K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_l - x}{h}\right) dx \\
&\quad \times \int D^2(p,\tau) K^2(\tau) d\tau \\
&= H_{11}^{(1,1)} + H_{11}^{(1,2)}, \text{ say.}
\end{aligned}
$$

It is straightforward to show $E\left|H_{11}^{(1,1)}\right| = O\left(n^{-1}h^{-3d_x/2}\right)$. Put $H_{11}^{(1,2)} = \frac{4}{n^2 h^{5d_x/2}} U_H$. Following analogous reasoning to the proof of Proposition A.8, we can show that U_H is a second order degenerate U-statistic satisfying $E|U_H|^2 = O(n^2 h^{3d_x})$. Thus, $E\left|H_{11}^{(1,2)}\right|^2 = O\left(n^{-2}h^{-2d_x}\right)$. Hence we have $H_{11}^{(1,1)} = o_P(1)$ and $H_{11}^{(1,2)} = o_P(1)$ by Markov's inequality and Chebyshev's inequality respectively. Following analogous reasoning to the proof of Lemma 3, we obtain the squared bias term $H_{11}^{(2)} = O\left(h^{-d_x/2+2p+2}\right) = o(1)$.

Now, we consider the H_{12} term. We decompose H_{12} as follows

$$
\begin{aligned}
H_{12} &= 2h^{-d_x/2} \iiint \left[|\phi_y(u,x)|^2 - 1\right] Re\left\{\left[\hat{\phi}_z(v,x) - \bar{\phi}_z(v,x)\right]\phi_z(v,x)^*\right\} dW_1(u) dW_2(v) a(x) dx \\
&\quad + 2h^{-d_x/2} \iiint \left[|\phi_y(u,x)|^2 - 1\right] Re\left\{\left[\bar{\phi}_z(v,x) - \phi_z(v,x)\right]\phi_z(v,x)^*\right\} dW_1(u) dW_2(v) a(x) dx \\
&= H_{12}^{(1)} + H_{12}^{(2)}, \text{ say.}
\end{aligned}
$$

Since $E\left(H_{12}^{(1)}\right)^2 = O\left(n^{-1}h^{-d_x}\right)$, we have $H_{12} = o_P(1)$ by Chebyshev's inequality. The bias term $H_{12}^{(2)} = O\left(h^{p+1-d_x/2}\right) = o(1)$. Thus, we have proved $H_1 = o_P(1)$. ■

Proof of Theorem 2. Under $\mathbb{H}_1(a_n)$, where $\sigma(u,v,x) = \phi_{yz}(u,v,x) - \phi_y(u,x)\phi_z(v,x) = a_n \delta(u,v,x)$, we can decompose

$$
\begin{aligned}
\hat{M}_h &= h^{d_x/2} \sum_{t=1}^{n} \iint \left|\hat{\sigma}(u,v,X_t)\right|^2 a(X_t) dW_1(u) dW_2(v) \\
&= h^{d_x/2} \sum_{t=1}^{n} \iint \left|\hat{\sigma}(u,v,X_t) - \sigma(u,v,X_t)\right|^2 a(X_t) dW_1(u) dW_2(v) \\
&\quad + 2 \sum_{t=1}^{n} \iint Re\left\{\left[\hat{\sigma}(u,v,X_t) - \sigma(u,v,X_t)\right]\sigma(u,v,X_t)^*\right\} a(X_t) dW_1(u) dW_2(v) \\
&\quad + h^{d_x/2} \sum_{t=1}^{n} \iint |\sigma(u,v,X_t)|^2 a(X_t) dW_1(u) dW_2(v) \\
&= h^{d_x/2} \sum_{t=1}^{n} \iint \left|\left(\hat{\phi}_{yz} - \phi_{yz}\right) - \phi_y\left(\hat{\phi}_z - \phi_z\right) - \phi_z\left(\hat{\phi}_y - \phi_y\right) - \left(\hat{\phi}_y - \phi_y\right)\left(\hat{\phi}_z - \phi_z\right)\right|^2 a(X_t) dW_1(u) dW_2(v)
\end{aligned}
$$

$$+ 2a_n h^{d_x/2} \sum_{t=1}^{n} \iint Re\{[(\hat{\phi}_{yz} - \phi_{yz}) - \phi_y(\hat{\phi}_z - \phi_z) - \phi_z(\hat{\phi}_y - \phi_y) - (\hat{\phi}_y - \phi_y)(\hat{\phi}_z - \phi_z)]\delta(u, v, X_t)^*\}$$
$$\times\, a(X_t) dW_1(u) dW_2(v)$$
$$+ a_n^2 h^{d_x/2} \sum_{t=1}^{n} \iint |\delta(u, v, X_t)|^2 a(X_t) dW_1(u) dW_2(v)$$
$$= \sum_{i=1}^{10} T_i + M_1 + M_2, \text{ say,}$$

where $\{T_i\}_{i=1}^{10}$ are defined as in $(A.2)$. Following the proof of Theorem 1, we can show that $\left(\sum_{i=1}^{10} T_i - B\right)/\sqrt{V} \xrightarrow{d} N(0,1)$ as $n \to \infty$ under $\mathbb{H}_1(a_n)$, where B and V are given by $(A.3)$ and $(A.4)$ respectively.

Now, we show $M_1 = o_P(1)$. We can decompose M_1 into four terms, denoted as $M_1^{(i)}, i = 1, \ldots, 4$, and show $M_1^{(i)} = o_P(1)$ for $i = 1, \ldots 4$. Since the proofs of the $M_1^{(i)}$ are similar, we focus on the proof of $M_1^{(1)} = o_P(1)$. We decompose

$$M_1^{(1)} = 2a_n h^{d_x/2} \sum_{t=1}^{n} \iint Re\{[\hat{\phi}_{yz}(u, v, X_t) - \bar{\phi}_{yz}(u, v, X_t)]\delta(u, v, X_t)^*\} a(X_t) dW_1(u) dW_2(v)$$
$$+ 2a_n h^{d_x/2} \sum_{t=1}^{n} \iint Re\{[\bar{\phi}_{yz}(u, v, X_t) - \phi_{yz}(u, v, X_t)]\delta(u, v, X_t)^*\} a(X_t) dW_1(u) dW_2(v)$$
$$= M_1^{(1,1)} + M_1^{(1,2)}, \text{ say.}$$

It is straightforward to show $E\left|M_1^{(1,1)}\right|^2 = O\left(h^{d_x/2}\right)$ and $E\left|M_1^{(1,2)}\right| = O\left(n^{1/2} h^{r+d_x/4}\right)$. Therefore, $M_1^{(1,1)} = o_P(1)$ and $M_1^{(1,2)} = o_P(1)$ by Chebyshev's inequality and Markov's inequality respectively. It follows that $\hat{M}^{(1)} = o_P(1)$. Similarly, we can also show $M_1^{(i)} = o_P(1)$ for $i = 2, 3, 4$. Therefore, $M_1 = o_P(1)$.

We now turn to M_2. By the weak law of large numbers, we have, as $n \to \infty$,

$$M_2 = n^{-1} \sum_{t=1}^{n} \iint |\delta(u, v, X_t)|^2 a(X_t) dW_1(u) dW_2(v) \xrightarrow{p} \gamma$$
$$= \iiint |\delta(u, v, x)|^2 dW_1(u) dW_2(v) a(x) g(x) dx.$$

In addition, under $\mathbb{H}_1(a_n)$, the asymptotic variance of $\hat{M}_h$, $\text{avar}(\hat{M}_h) = \text{avar}\left(\sum_{i=1}^{10} T_i\right) = V$ given $M_1 = o_P(1)$ and $M_2 - \gamma = o_P(1)$. Consequently, we obtain the desired result of Theorem 2. ■

Proof of Theorem 3. Similar to the proof of Theorem 2. ■

https://doi.org/10.1017/S026646661700010X Published online by Cambridge University Press