Characteristic Function Based Testing for Conditional Independence: A Nonparametric Regression Approach

Xia Wang[a], Yongmiao Hong[b,*]

[a] Wang Yanan Institute for Studies in Economics (WISE) and MOE Key Laboratory of Econometrics, Xiamen University

[b] Department of Economics and Department of Statistical Sciences, Cornell University

November, 2012

Abstract

We propose a characteristic function based test for conditional independence, which is applicable in both cross-sectional and time series contexts. Our test is not only asymptotically locally more powerful than Su and White's (2007) test, but also is more flexible in inferring patterns of conditional dependence. In addition to our ominous test, we also propose a class of derivative tests to gauge possible patterns of conditional dependence. These derivative tests deliver some interesting model-free tests for such important hypotheses as omitted variables, Granger causality in mean and conditional uncorrelatedness. All proposed tests have a convenient asymptotic $N(0,1)$ distribution under the null hypotheses. Unlike many other smoothed nonparametric tests for conditional independence, we allow nonparametric estimators for both conditional joint and marginal characteristic functions to jointly determine the asymptotic distribution of our tests. This leads to a much better size performance in finite samples. Monte Carlo studies demonstrate the well behavior of our tests in finite samples. In an application to testing nonlinear Granger causality, we document the existence of nonlinear relationships between money and output, which may be ignored by the linear Granger causality test and Su and White's (2007) test.

Keywords: Characteristic function, Conditional independence, Conditional Uncorrelatedness, Curse of dimensionality, Granger causality, Local linear estimation, Local power, Money and output, Omitted variable test

JEL classification: C12, C14.

---

*Correspondence: Yongmiao Hong, Ernest S. Liu Professor of Economics and International Studies, Department of Economics and Department of Statistical Sciences, Cornell University, 424 Uris Hall, Ithaca, NY, 14850, USA; Email: yh20@cornell.edu. This project was started when Wang was an exchange doctoral student to Department of Economics, Cornell University, in Spring 2012. We thank Liangjun Su for kindly providing his computer codes of his test for conditional independence. All remaining errors are solely our own.

1 Introduction

Conditional independence is an important concept in probability theory and a widely used assumption in economic and financial modeling. Let X, Y and Z be random variables or vectors. Then Y is independent of Z given X, which we denote as $Y \perp Z | X$, if the joint probability density function of Y and Z conditional on X equals to the product of the conditional marginal density functions of Y and Z for any values of (Y, Z) in their support. The conditional independence assumption encompasses many important hypotheses in econometrics and statistics (Dawid, 1979; 1980). To motivate the important roles of the conditional independence assumption, we provide a few examples in economics and econometrics.

The first example of conditional independence is the Markov property of a time series process. A strictly stationary time series X_t is said to follow a Markov process if

$$X_{t+1} \perp (X_{t-1}, X_{t-2}, \cdots) | X_t$$

For a Markov process $\{X_t\}$, the current state variable or vector X_t will contain all useful information in predicting the future behavior of X_t. The Markov property is broadly used in economics and finance (e.g., Easley and O'Hara, 1987; Rust, 1994). It is a fundamental property in time series analysis and is widely accepted in econometric testing (Bouissou et al., 1986; Aït-Sahalia et al., 2009) and economic modeling (Beja, 1979; Pakes and McGuire, 2001). In particular, if the Markov property holds, we can reduce the time dimension and capture a large amount of data information using a simple time series model with only one lag. However, if this property is violated, then economic models and predictions based on the Markov assumption will be suboptimal. Recently, such literature as Aït-Sahalia et al. (2010), Chen and Hong (2012) propose some nonparametric tests for Markov Property, which are quite useful in testing Markov hypothesis in practice.

The second example of conditional independence is non-Granger causality, which was first proposed by Granger (1969, 1980). Given two time series processes $\{Z_t\}$ and $\{Y_t\}$, and lag orders p and q, $\{Z_t\}$ does not Granger cause $\{Y_t\}$ in distribution if

$$Y_t \perp Z_{t-q}^{t-1} | Y_{t-p}^{t-1}$$

where $Z_{t-q}^{t-1} \triangleq \{Z_{t-1}, Z_{t-2}, \cdots, Z_{t-q}\}$, $Y_{t-p}^{t-1} \triangleq \{Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}\}$. If the hypothesis of distributional non-Granger causality is rejected, then the information Z_{t-q}^{t-1} is useful in predicting the future distribution of $\{X_t\}$. Granger (1969) proposes a convenient F test for Granger causality in a linear regression framework, which is a special case of Granger causality in mean. Considerable empirical studies in the literature have focused on testing linear Granger causality between financial and economic variables using Granger's (1969) F test, such as Thornton and Batten (1985), Granger et al. (2000), Calderón and Liu (2003). Although they find some interesting linear economic relationships, they may miss some important nonlinear phenomenons, such as the asymmetric effect of monetary policy (Kim and Nelson, 2006) and asymmetric behavior of financial returns (Campbell, 1992; Peiró, 1999).

The third example of conditional independence is the missing at random assumption, which is widely maintained in treatment response analysis and missing data problems, such as Hahn et al. (2001) and

Wang et al. (2004). The data is missing at random if the missingness does not depend on the values of variables in the data set subject to analysis (Rubin, 1976). More specifically, for a triple of random variables $\{X, Y, Z\}$, where Y is an outcome variable, X is an explanatory variable, and Z is a binary variable indicating treatment, which is equal to 1 if Y is observed and 0 otherwise. Here, a researcher observes (X, Z) but observes Y only when $Z = 1$. The variable Y is missing at random conditional on X if:

$$Y \perp Z | X.$$

In empirical analysis, if the missing at random assumption holds, one can obtain consistent estimation by simply throwing away the unobservable samples and point-identify treatment effect in response analysis. However, abuse of this assumption generally yields inconsistent estimation, which is called selectivity bias in the literature (Heckman, 1976; Little, 1985). In addition, Horowitz and Manski (2000) and Manski (2000, 2003, 2007) show that, absent the ignorable treatment selection assumption, we can only obtain interval estimators for the treatment effect and mean response function rather than point-identify them.

The last example of conditional independence is exogeneity. Suppose the random variable Y is generated by the following unknown structural function:

$$Y = g(X, U) \tag{1}$$

where X is an observed explanatory variable or vector, and U is an unobserved error term. In the literature, X is called an exogenous variable in Eq.(1) if $X \perp U$. To test exogeneity, some researchers (e.g., Blundell and Horowitz, 2007; Lee, 2009) 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 theory and economic modeling. Generally, econometric models suffer from endogenous problem usually require different estimation approaches, which are always less efficient than those when all explanatory variables are exogenous. For example, Hall and Horowitz (2005) document the estimation efficiency loss due to the use of instrumental variables when explanatory variables are endogenous in nonparametric regression analysis of Eq.(1) .

Motivated by the widespread applications of the conditional independence assumption, a growing literature focuses on testing conditional independence. Linton and Gozalo (1996) test the conditional independence assumption for an i.i.d. data generation process (DGP) using the empirical distribution function. Su and White (2007, 2008) develop nonparametric tests based on the weighted distances between the conditional characteristic functions and between the densities respectively. Huang (2010) tests conditional independence based on a maximal nonlinear conditional correlation. Su and White (2011a) propose two smoothed empirical likelihood ratio tests by representing the null hypothesis as an infinite collection of conditional moment restrictions. Su and White (2011b) test conditional independence using local polynomial quantile regression which has the appealing advantage of parametric convergence rate, but with the cost of a non-pivotal asymptotic distribution. Bouezmarni et al. (2012) develop a nonparametric copula-based test for conditional independence.

In this paper, we propose a characteristic function-based test for conditional independence using a nonparametric regression approach. Compared with Su and White (2007) and other related nonparametric tests for conditional independence in the literature, our test has the following features:

First of all, our test can detect a class of local alternatives that converge to the null hypothesis of conditional independence at a faster convergence rate than Su and White's (2007) test and some other tests in the literature. Suppose the dimensions of random vectors X, Y, Z are d_x, d_y, d_z respectively and we are interested in testing whether Y is independent of Z conditional on X. Let n and $h = h(n)$ denote the sample size and the bandwidth used in our nonparametric test. Then, thanks to the use of a regression approach and conditional characteristic functions, the convergence rate of the class of local alternatives for our test is $n^{-1/2}h^{-d_x/4}$, which is faster than the rate of local alternatives for most nonparametric tests aforementioned, including Su and White (2007, 2008, 2011a), Bouezmarni et al. (2012), which depends not only on d_x but also the dimensions of other variables. Since the convergence rate of our test only depends on the dimension of X, it is less severely subjected to the notorious "curse of dimensionality" problem than the aforementioned tests.

Secondly, our test is more flexible in gauging possible sources of conditional dependence. As is well known, the characteristic function can be differentiated to obtain various moments. By differentiating our ominous test statistic with respect to auxiliary parameters at the origin up to various orders, we obtain a class of useful derivative tests, including tests of omitted variables, Granger causality in mean, and conditional uncorrelatedness. In addition, all of these tests have a convenient asymptotic one-side $N(0, 1)$ distribution under the null hypothesis.

Thirdly, unlike many nonparametric tests of conditional independence, we use a single bandwidth rather than two different bandwidths in estimating both the conditional joint and marginal characteristic functions, which significantly improves the size performance of the proposed test in finite samples and avoids difficulties in choosing multi-bandwidths. In the most of the previous literature, different bandwidths are used when estimating joint and marginal densities (or joint and marginal characteristic functions). In particular, certain conditions on the relative speeds of the bandwidths are imposed, so that nonparametric estimators of marginal densities (or marginal characteristic functions) have faster convergence rates and thus have no impact on the asymptotic distribution of the proposed tests, although the order of magnitude for marginal density estimators may be rather close to that of joint density estimators. In contrast, since we choose a common bandwidth, nonparametric estimation errors of the conditional joint and marginal characteristic functions jointly affect the asymptotic distribution of our test statistic. This renders it more challenging to derive the asymptotic distribution of our test. However, it is expected to result in a better size of the test in finite samples due to fewer negligible high order terms. In addition, unlike Su and White's (2007) requirement of using a high order kernel function, we allow to use positive (i.e., second order) kernel functions. Besides, we apply local linear regression to estimate the conditional characteristic functions, which has significant advantages over the Nadaraya-Watson kernel estimator, particularly in reducing the bias, and adapting automatically to the boundary bias due to asymmetric coverage of data in the boundary regions.

Finally, our tests are applicable to both cross-sectional and time series data. The proposed tests follow a convenient asymptotic $N(0,1)$ distribution under the null hypotheses. Furthermore, while we require the conditioning variable X to be a continuous random variable, we allow both Y and Z to be either discrete or continuous random variables or a mixture of them.

The paper is organized as follows. In Section 2, we formalize the conditional independence condition and state the hypothesis of interest. In Section 3, we propose a characteristic function based test for conditional independence using a nonparametric regression approach. We derive the asymptotic null distribution of our test statistic in Section 4 and investigate its asymptotic local power property in Section 5. In Section 6, we develop a class of derivative tests to gauge possible patterns of conditional dependence by differentiating our ominous test with respect to auxiliary parameters at the origin. In Section 7, we study the finite sample performance of our test in comparison with Su and White's (2007) test. Section 8 considers an empirical application to nonlinear Granger causality between money and output. Conclusions are provided in Section 9. All proofs are relegated to the Appendix.

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 n identical distributed but weakly dependent observations $(X_t, Y_t, Z_t), t = 1, 2, \cdots, n$. As our results are applicable in both cross-sectional and time series contexts, the index t stands for time in a time series context, and it denotes a cross-sectional unit (e.g., household, firm, etc) in a cross-sectional context. Denote $f(\cdot|\cdot)$ as the conditional density (mass) function of one random vector given another. For convenience, the function $f(\cdot|\cdot)$ is referred to as conditional density function below. However, we allow both Y and Z to be either discrete or continuous random variables or a mixture of them and our results are same for each case. Our null hypothesis of interest is that conditional on X, the random vectors Y and Z are independent, i.e.

$$\mathbb{H}_0 : P\left[f(y, z|X) = f(y|X)f(z|X)\right] = 1 \text{ for any } (y, z) \in \mathbb{R}^{d_y + d_z}. \tag{2}$$

The alternative hypothesis is

$$\mathbb{H}_A : P\left[f(y, z|X) \neq f(y|X)f(z|X)\right] > 0 \text{ for some } (y, z) \in \mathbb{R}^{d_y + d_z} \text{ with positive Lebesgue measure }. \tag{3}$$

Su and White (2007, 2008, 2011a) test whether Y and Z are independent conditional on X under the null hypothesis $P[f(y|X, Z) = f(y|X)] = 1$ for any $y \in \mathbb{R}^{d_y}$, which is equivalent to our null hypothesis (2). However, (2) only relates to densities conditional on X, which will only involve d_x dimensional smoothing when estimating conditional characteristic functions, whereas Su and White's (2007, 2008, 2011a) tests involve at least $d_x + d_z$ dimensional smoothing because they use the density of Y conditional on both X and Z. This, together with the use of characteristic functions, makes our test more powerful than Su and White's (2007) test and many other nonparametric tests. See Section 5 for more discussion.

As the Fourier transform of the conditional density, the conditional characteristic function can equally capture the entire conditional probability distribution of a random vector. Thus, we can also represent the null hypothesis using conditional characteristic functions. Denote:

$$\begin{aligned}\phi_{yz}(u,v,x) &= E\left(e^{\mathbf{i}(u'Y_t+v'Z_t)}|X_t=x\right),\\ \phi_y(u,x) &= E\left(e^{\mathbf{i}u'Y_t}|X_t=x\right),\\ \phi_z(v,x) &= E\left(e^{\mathbf{i}v'Z_t}|X_t=x\right).\end{aligned}$$

Furthermore, define a generalized conditional covariance

$$\sigma(u,v,x)=\text{cov}\left(e^{\mathbf{i}u'Y_t},e^{\mathbf{i}v'Z_t}|X_t=x\right),\quad (u,v)\in\mathbb{R}^{d_x+d_y}. \tag{4}$$

Straightforward algebra shows that

$$\sigma(u,v,X_t)=\phi_{yz}(u,v,X_t)-\phi_y(u,X_t)\phi_z(v,X_t). \tag{5}$$

Thus, $\sigma(u,v,X_t)=0$ for all $u,v\in\mathbb{R}^{d_y+d_z}$ with probability one if and only if Y_t and Z_t are independent conditional on X_t.

For a weakly stationary time series $\{Y_t\}$, when $Z_t=Y_{t-k}$ and $X_t=(Y_{t-1},\cdots,Y_{t-k+1})'$, $\sigma_k(u,v,x)$ could be regarded as a generalized partial autocovariance function which is similar as Hong's (1999) generalized autocovariance function. It can capture any type of pairwise conditional partial dependence over various lags, including those with zero partial autocovariance, such as the bilinear, nonlinear moving average, ARCH/GARCH processes. To see this, we rewrite $\sigma_k(u,v,x)$ using the Taylor series expansion:

$$\begin{aligned}\sigma_k(u,v,x) &= \text{cov}\left(\sum_{m=0}^{\infty}\frac{(\mathbf{i}uY_t)^m}{m!},\sum_{l=0}^{\infty}\frac{(\mathbf{i}vY_{t-|k|})^l}{l!}|X_t=x\right)\\ &= \sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\frac{(\mathbf{i}u)^m(\mathbf{i}v)^l}{m!l!}\text{cov}\left(Y_t^m,Y_{t-|k|}^l|X_t=x\right)\end{aligned}$$

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

Using the definition of $\sigma(u,v,x)$, the hypotheses $\mathbb{H}_0$ and $\mathbb{H}_A$ can be equivalently represented as follows:

$$\mathbb{H}_0:\quad P\left[\sigma(u,v,X_t)=0\right]=1 \text{ for any } (u,v)\in\mathbb{R}^{d_y+d_z} \tag{6}$$

versus

$$\mathbb{H}_A:\quad P\left[\sigma(u,v,X_t)\neq 0\right]>0 \text{ for some } (u,v)\in\mathbb{R}^{d_y+d_z}. \tag{7}$$

It is important to emphasize that we must check (6) for all the $(u,v)\in\mathbb{R}^{d_y+d_z}$ rather than only a subset of $\mathbb{R}^{d_y+d_z}$. Although this is quite involved, it offers some appealing features of our test. For example, by differentiating the generalized conditional covariance function $\sigma(u,v,X_t)$ with respect to u or v or both at the origin, we can infer possible patterns of conditional dependence, which may provide valuable information for modeling economic relationships.

3 Nonparametric Regression Based Testing

3.1 Generalized Nonparametric Regression

Recall that $\sigma(u, v, X_t) = 0$ a.s. for all $(u, v) \in \mathbb{R}^{d_y+d_z}$ under $\mathbb{H}_0$. Given an observed sample $\{X_t, Y_t, Z_t\}_{t=1}^n$ of size n, we shall estimate the generalized conditional covariance $\sigma(u, v, X_t)$ nonparametrically and check whether it is identically zero for all $(u, v) \in \mathbb{R}^{d_y+d_z}$. We shall estimate $\phi_{yz}(u, v, x)$, $\phi_y(u, x)$ and $\phi_z(v, x)$ nonparametrically, which are potentially highly nonlinear. Since $\phi_{yz}(u, v, x)$, $\phi_y(u, x)$ and $\phi_z(v, x)$ are generalized regression functions, namely, $\phi_{yz}(u, v, x) = E(e^{\mathbf{i}(u'Y_t+v'Z_t)}|X_t = x)$, $\phi_y(u, x) = \phi_{yz}(u, 0, x)$, and $\phi_z(v, x) = \phi_{yz}(0, v, x)$, we use local linear regression. Compared with the Nadaraya-Watson estimator, the local linear regression estimator can not only reduce bias in the interior region, but also automatically corrects boundary biases. Since the conditional marginal characteristic functions $\phi_y(u, x)$ and $\phi_z(v, x)$ can be obtained from the conditional joint characteristic function $\phi_{yz}(u, v, x)$ by setting $v = 0$ or $u = 0$ respectively, we only need construct a nonparametric estimator for $\phi_{yz}(u, v, x)$.

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

$$\min_{\beta\in\mathbb{C}^{d_x+1}} \sum_{t=1}^{n} |e^{\mathbf{i}(u'Y_t+v'Z_t)} - \beta_0 - \beta_1'(X_t - x)|^2 K_h(X_t - x), \text{ where } x \in \mathbb{R}^{d_x}, u \in \mathbb{R}^{d_y}, v \in \mathbb{R}^{d_z}, \quad (8)$$

where $\beta = (\beta_0, \beta_1')'$ is a $(d_x + 1) \times 1$ parameter vector, $K_h(x) = h^{-d_x}K(\frac{x}{h})$, and $K : \mathbb{R}^{d_x} \to \mathbb{R}$ is a kernel function, and $h = h(n)$ is a bandwidth. The solution to the problem (8) is:

$$\hat{\beta} \equiv \hat{\beta}(u, v, x) = (X'WX)^{-1}X'WV, \quad x \in \mathbb{R}^{d_x}, \quad (9)$$

where X is a $n \times (d_x + 1)$ matrix with the ith row given by $[1, (X_i - x)']$, $W = \text{diag}[K_h(X_1 - x), \cdots, K_h(X_n - x)]$, and $V = \left(e^{\mathbf{i}(u'Y_1+v'Z_1)}, \cdots, e^{\mathbf{i}(u'Y_n+v'Z_n)}\right)'$.

The conditional expectation $\phi(u, v, x) = E\left(e^{\mathbf{i}(u'Y_t+v'Z_t)}|X_t = x\right)$ can be estimated by the local intercept estimator $\hat{\beta}_0(u, v, x)$, that is:

$$\hat{\phi}_{yz}(u, v, x) = \sum_{t=1}^{n} \hat{W}\left(\frac{X_t - x}{h}\right) e^{\mathbf{i}(u'Y_t+v'Z_t)}, \quad (10)$$

where $\hat{W}(\cdot)$ is an effective kernel, defined as

$$\hat{W}(t) \equiv e_1'S_n^{-1}\,[1\; th]'K(t)/h^{d_x}, \quad (11)$$

$e_1 = (1, 0, \cdots, 0)'$ is a $(d_x+1)\times 1$ unit vector, $S_n = X'WX$ is a $(d_x+1)\times(d_x+1)$ matrix. According to Hjellvik et al. (1998) and Chen and Hong (2010), the effective kernel can be written as:

$$\hat{W}(t) = \frac{1}{nh^{d_x}g(x)}K(t)[1 + o_P(1)]. \quad (12)$$

By setting $v = 0$ or $u = 0$ respectively, we can obtain the estimators of the conditional marginal characteristic functions $\phi_y(u, x)$ and $\phi_z(v, x)$:

$$\hat{\phi}_y(u, x) = \hat{\phi}_{yz}(u, 0, x) = \sum_{t=1}^{n} \hat{W}\left(\frac{X_t - x}{h}\right) e^{\mathbf{i}u'Y_t}, \quad (13)$$

$$\hat{\phi}_z(v, x) = \hat{\phi}_{yz}(0, v, x) = \sum_{t=1}^{n} \hat{W}\left(\frac{X_t - x}{h}\right) e^{\mathbf{i}v'Z_t}. \quad (14)$$

We note that the nonparametric estimators for all three conditional expectations, $\phi_{yz}(u, v, x)$, $\phi_y(u, x)$, $\phi_z(v, x)$, only involves smoothing over the conditional variable X_t. This differs from nonparametric estimators for conditional densities such as $f(y, z|x)$ or $f(y|z, x)$, which will involve smoothing over Y_t, Z_t, X_t simultaneously. It also differs from nonparametric estimators for the conditional characteristic function $E(e^{\mathbf{i}u'Y_t}|X_t, Z_t)$, which involves smoothing over X_t and Z_t simultaneously. The reduction in curse of dimensionality renders our test more powerful than most existing nonparametric tests of conditional independence. See Section 5 for asymptotic local power analysis.

3.2 Nonparametric Based Test Statistic

Under the null hypothesis, we have $\sigma(u, v, X_t) = 0\ a.s.$ for all $u,\ v$. Therefore, we can measure

$$\mathbb{H}_0 : \hat{\sigma}(u, v, X_t) = \hat{\phi}_{yz}(u, v, X_t) - \hat{\phi}_y(u, X_t)\hat{\phi}_z(v, X_t) = 0$$

via the quadratic form:

$$\hat{M} = \frac{1}{n} \sum_{t=1}^{n} \iint |\hat{\sigma}(u, v, X_t)|^2 a(X_t) dW_1(u) dW_2(v) \quad (15)$$

where $a : \mathbb{R}^{d_x} \to \mathbb{R}^+$ is a weighting function for the conditioning vector X_t, $W_1 : \mathbb{R}^{d_y} \to \mathbb{R}^+$ and $W_2 : \mathbb{R}^{d_z} \to \mathbb{R}^+$ are nondecreasing weighting functions of u, v that weight supports symmetric about the origin equally. The weighting function $a(\cdot)$ is commonly used in the literature to truncate integrations (e.g., Hjellvik et al., 1998; Aït-Sahalia et al., 2001; Chen and Hong, 2010). Since nonparametric estimation at sparse extreme observations is inaccurate in finite samples, by choosing an appropriate weighting function $a(\cdot)$, one can alleviate the influences of unreliable estimates. The introduction of weighting functions $W_1(u)$ and $W_2(v)$ allows us to consider many points for u, v. One popular weighting function in the literature is the $N(0, 1)$ cumulative distribution function (CDF). In fact, W_1 and W_2 need not to be continuous functions. Any nondecreasing function with countable discontinuity points satisfies our requirement for W_1, W_2. One special case of discontinuous weighting function is a discrete multivariate CDF, which allows us to consider a countable number of grid points of (u, v) and provides a convenient way to avoid high dimensional integration.

Our test statistic (15) has the following standardized version:

$$\widehat{SM} = \left[h^{d_x/2} \sum_{t=1}^{n} \iint |\hat{\sigma}(u, v, X_t)|^2 a(X_t) dW_1(u) dW_2(v) - \hat{B} \right] / \sqrt{\hat{V}} \quad (16)$$

where

$$\hat{B} = h^{-d_x/2}\int\left[\int\left(1-|\hat{\phi}_y(u,x)|^2\right)dW_1(u)\right]\left[\int\left(1-|\hat{\phi}_z(v,x)|^2\right)dW_2(v)\right]a(x)dx \times\int K^2(\tau)d\tau, \quad (17)$$

$$\hat{V} = 2\int\left[\iint|\hat{\Phi}_y(u_1+u_2,x)|^2dW_1(u_1)dW_1(u_2)\iint|\hat{\Phi}_z(v_1+v_2,x)|^2dW_2(v_1)dW_2(v_2)\right]a^2(x)dx \times\int\left[\int K(\tau)K(\tau+\eta)d\tau\right]^2d\eta, \quad (18)$$

$\hat{\Phi}_s(a_1+a_2,x) = \hat{\phi}_s(a_1+a_2,x) - \hat{\phi}_s(a_1,x)\hat{\phi}_s(a_2,x)$, and $\hat{\phi}_s(a,x)$ is the nonparametric estimator for $\phi_s(a,x)$, $s = y$ or z.

The factors $\hat{B}$ and $\hat{V}$ are the estimators for the asymptotic mean and variance of the quadratic form in Eq.(15). The asymptotic variance estimator $\hat{V}$ involves $2\max\{d_y, d_z\} + d_x$ dimensional integration. When the dimensions of X, Y or Z are high, the calculation of $\widehat{SM}$ depends on high-dimensional integration. In practice, one can adopt numerical integration or simulation techniques. Alternatively, one can simply use a finite number of grid points for (u, v) or refer to other technical integration methods, such as sequential Monte Carlo simulation.

Both $\hat{B}$ and $\hat{V}$ in (17) and (18) are derived under the null hypothesis as the sample size $n \to \infty$. However, they may not approximate well the mean and variance of the statistic (15) in finite samples respectively, which may lead to poor size. To improve the size of the test in finite samples, we also consider the following finite-sample version test statistic:

$$\widehat{SM}_n = \left[h^{d_x/2}\sum_{t=1}^{n}\iint|\hat{\sigma}(u,v,X_t)|^2a(X_t)dW_1(u)dW_2(v) - \hat{B}_n\right]/\sqrt{\hat{V}} \quad (19)$$

where

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

with $\hat{\varepsilon}_y(u, X_s) = e^{iu'Y_s} - \hat{\phi}_y(u, X_s)$ and similarly for $\hat{\varepsilon}_z(v, X_s)$. The complex-valued random variables $\hat{\varepsilon}_y(u, X_s)$ and $\hat{\varepsilon}_z(v, X_s)$ could be viewed as estimated generalized regression errors. One could also replace the scaling factor $\hat{V}$ by its finite-sample version

$$\hat{V}_n = 2h^{d_x/2}\sum_{1\le r<s\le n}\left\{\sum_{t=1}^{n}a(X_t)\hat{W}\left(\frac{X_s-X_t}{h}\right)\hat{W}\left(\frac{X_r-X_t}{h}\right) \times\iint Re\left[(\hat{\varepsilon}_y(u,X_s)\hat{\varepsilon}_z(v,X_s))(\hat{\varepsilon}_y(u,X_r)\hat{\varepsilon}_z(v,X_r))^*\right]dW_1(u)dW_2(v)\right\}^2,$$

where $Re(A)$ and A^* denote the real part and the conjugate of a complex-valued number A respectively. However, since it is tedious and computational costly when n is large, we do not intend to do so. Simulation studies show that the test statistics in Eq.(19) works reasonably well in finite samples.

4 Asymptotic Distribution

In this section, we will derive the asymptotic distribution of the omnibus test statistic $\widehat{SM}$ under the null hypothesis. We first impose the following regularity conditions.

Assumption A.1 [*Data Generating Process*]: Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. (a) The stochastic vector process $W_t \equiv (X_t', Y_t', Z_t')'$, $t = 1, \cdots, n$ is a strictly stationary absolutely regular process 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 function $g(x)$ of X_t is positive, bounded, continuous, and twice differentiable for all $x \in G$, where G is a compact support set of X_t in $\mathbb{R}^{d_x}$.

Assumption A.2 [*Conditional Characteristic Function*]: Let $\phi_{yz}(u, v, x),\ \phi_y(u, x),\ \phi_z(v, x)$ be the conditional characteristic functions of $(Y_t, Z_t),\ Y_t,\ Z_t$ given $X_t = x$ respectively. For each $u \in \mathbb{R}^{d_y},\ v \in \mathbb{R}^{d_z}$, $\phi_{yz}(u, v, x)$, $\phi_y(u, x)$, and $\phi_z(v, x)$ are measurable and twice continuously differentiable with respect to $x \in G$.

Assumption A.3 [*Kernel Function*]: $K : \mathbb{R}^{d_x} \to \mathbb{R}^+$ is a product kernel function 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 a symmetric, bounded, and twice continuously differentiable function with $\int_{-\infty}^{\infty} k(u)du = 1,\ \int_{-\infty}^{\infty} uk(u)du = 0$ and $\int_{-\infty}^{\infty} u^2 k(u)du = C_k < \infty$.

Assumption A.4 [*Weighting Functions*]: (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 weight sets symmetric about zero equally, with $\int_{\mathbb{R}^{d_y}} \|u\|^4 dW_1(u) < \infty,\ \int_{\mathbb{R}^{d_z}} \|v\|^4 dW_2(v) < \infty$; (b) $a : G \to \mathbb{R}^+$ is a bounded weighting function that is continuous over G, where $G \subset \mathbb{R}^{d_x}$ is a compact support set of X_t given in Assumption A.1.

Assumption A.1 imposes regularity conditions on the DGP. Assumption A.1(a) is standard for application of a central limit theorem for U statistics for weakly dependent data (e.g. Hjellvik et al., 1998). The β-mixing condition restricts the degree of temporal dependence in (X_t, Y_t, Z_t), which is generally adopted in the nonparametric time series literature, 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 autoregressive moving average (ARMA), bilinear, and autoregressive conditional heteroscedastic (ARCH) process, satisfy the β-mixing condition (Fan and Li, 1999). Assumption A.1(b) is a smoothness condition, which rules out discrete random variables for X_t. However, we could extend our test to incorporate the discrete case of X_t in a similar way to Su and White (2008). Note that we allow the components of Y_t and Z_t to be either continuous or discrete random variables or a mixture of them.

Assumption A.2 provides conditions on conditional characteristic functions, which are the Fourier transforms of conditional density functions. We can easily translate these conditions into the conditions on the conditional density functions (when they exist). In particular, Assumption A.2 holds if $f(Y_t, Z_t|x),\ f(Y_t|x),\ f(Z_t|x)$ are measurable and twice continuously differentiable with respect to $x \in G$ with probability one.

Assumption A.3 allows the use of familiar positive kernels, such as the Gaussian and Epanechnikov kernels. Unlike Su and White (2007, 2008), we do not have to use a higher order kernel, which could reduce biases to a higher order but at the cost of a larger variance, and therefore would affect the asymptotic efficiency of our test.

Assumption A.4 imposes some mild conditions on the weighting functions $W_1(u), W_2(v)$ and $a(x)$, respectively. These conditions ensure the existence of the integral in (16). Many functions satisfy the conditions for $W_1(u)$ and $W_2(v)$, an example being the CDFs with finite fourth order moment. In addition, $W_1(u)$ and $W_2(v)$ need not be continuous, which allows us to compute our test statistic using a finite number of grid points of (u, v) instead of computational costly numerical integration or simulation methods. For convenience, we can use the weighting functions of multiplicative form:

$$W_1(u) = \prod_{i=1}^{d_y} w(u_i) \text{ and } W_2(v) = \prod_{i=1}^{d_z} w(v_i),$$

where $w : \mathbb{R} \to \mathbb{R}^+$ 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, $h = cn^{-\lambda}$ for $1 \leq d_x \leq 8$, and $\frac{1}{d_x+4} < \lambda < \frac{2}{3d_x}$, where $0 < c < \infty$. Then $\widehat{SM} \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$ as $n \to \infty$.*

The proof relies on the central limit theorem for degenerate U-statistics for weakly dependent processes in Tenreiro (1997), which has been widely used by Su and White (2007, 2008) and Hong and Lee (2013). To derive the asymptotic distribution, we decompose $\hat{M}$ into the sum of ten terms, from which, we obtain six U-statistics. These six U-statistics jointly determine the asymptotic distribution of our test statistic. Since the proof is quite involved, we relegate it to Appendix.

Theorem 1 restricts the dimension of X_t no more than 8. This condition is not severely restrictive in practice. For comparison, Su and White (2008) restrict the sum of the dimensions of X_t, Y_t, Z_t to be no more than 7. Moreover, we do not impose any constraints on the dimensions of Y_t and Z_t because the convergence rate of our test statistic only depends on the dimension of X_t. This differs from Su and White (2007, 2008), which involve smoothing of $d_x + d_z$ and $d_x + d_y + d_z$ dimensions respectively, and therefore require restrictions on the sum of the dimensions of X_t, Y_t, Z_t.

Theorem 1 allows the choice of a wide range of admissible rates for the bandwidth h. In order to reduce the number of leading terms in asymptotic mean and variance, and to avoid estimating the Laplacian of conditional characteristic functions, we rule out the optimal bandwidth that minimizes integrated mean squared error (IMSE). Since our assumption in Theorem 1 requires $nh^{k+4} \to 0$, only the square of sampling error terms, e.g. $|\hat{\phi}_{yz}(u,v,x) - E\hat{\phi}_{yz}(u,v,x)|^2$, $Re[(\hat{\phi}_{yz}(u,v,x) - E\hat{\phi}_{yz}(u,v,x))(\hat{\phi}_y(u,x) - E\hat{\phi}_y(u,x))^*]$, affect the limiting distribution of our test statistic.

It is important to emphasize that we adopt the same bandwidth h in estimating the conditional joint characteristic function $\phi_{yz}(u,v,x)$ and the conditional marginal characteristic functions $\phi_y(u,x)$ and $\phi_z(v,x)$. As a result, both the nonparametric estimation errors from the conditional joint and marginal characteristic functions are of the same order of magnitude and they jointly determine the limiting distribution of our test statistic. This differs from most of the existing literature, where different bandwidths are used to estimate joint and marginal densities (or joint and marginal characteristic functions), and the relative speeds for bandwidths are carefully imposed so that the nonparametric estimators of marginal densities or marginal characteristic functions converge faster than their multivariate counterparts and thus they have no impact on the asymptotic distribution of the test statistic. This is the case approach taken by Fan and Li (1996), Lavergne and Vuong (2000), Aït-Sahalia et al. (2001), Su and Ullah (2009), Su and White (2007, 2011a). However, although the estimation errors for the marginal densities are higher order terms of that for the joint density, their magnitudes may be rather close to each other in finite samples. Due to the impact of neglecting higher order estimation errors from the marginal densities, the size of their tests may be poor in finite samples. In contrast, by choosing the same bandwidth, we expect that our approach will provide a better size performance in finite samples, as is confirmed in our simulation study. We also avoid the delicate business of choosing multi-bandwidths and do not have to choose a higher order kernel.

Our test is applicable in both cross-sectional and time series contexts. Under the null hypothesis, it is asymptotically pivotal and has a convenient asymptotic $N(0, 1)$ distribution. Hence, we can compare the test statistic $\widehat{SM}$ with the one-sided critical value z_α at significance level α from the $N(0, 1)$ distribution, and reject $\mathbb{H}_0$ when $\widehat{SM} > z_\alpha$. For example, the asymptotic critical value at the 5% significant level is 1.645. In contrast with Su and White's (2011b) test, our test is not only pivotal for independent or martingale difference sequence observations, but also has an asymptotic $N(0, 1)$ distribution for observations with weak dependence. Furthermore, in a time series context, since $\sigma(u, v, x)$ can be regarded as the generalized partial autocovariance function, our test is suitable for testing partial serial dependence and is powerful in detecting any type of partial serial dependence, including those with zero partial autocorrelation.

5 Asymptotic Local Power

Since both Su and White's (2007) test and our test are based on the characteristic function, it is interesting to compare their relative efficiency. We first consider the following class of local alternatives:

$$\mathbb{H}_1(a_n) : f(y, z|x) = f(y|x)f(z|x) + a_n q_a(y, z|x) \tag{20}$$

where $q_a(y, z|x)$ is a twice continuously differentiable function, which satisfies $q_a(y, z|x) \neq 0$ and $\iint q_a(y, z|x)dydz = 0$. The additional term $a_n q_a(y, z|x)$ characterizes the departure of the conditional joint density function from the product of conditional marginal density functions and the rate a_n is the speed at which the deviation vanishes to 0 as the sample size $n \to \infty$. By taking the Fourier transform of Eq. (20), we obtain

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

where $\delta(u, v, x) = \iint e^{\mathbf{i}(u'y+v'z)} q_a(y, z|x)dydz$ is the Fourier transform of $q_a(y, z|x)$ and satisfies:

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

Theorem 2. *Suppose Assumptions A.1-A.4 and* $\mathbb{H}_1(a_n)$ *hold with* $a_n = n^{-1/2}h^{-d_x/4}$*, and the bandwidth* $h = cn^{-\lambda}$ *for* $1 \leq d_x \leq 8$ *and* $\frac{1}{d_x+4} < \lambda < \frac{2}{3d_x}$*, where* $0 < c < \infty$*. Then, as* $n \to \infty$*, the power of the test satisfies*

$$P\left[\widehat{SM} \geq z_\alpha | \mathbb{H}_1(a_n)\right] \to 1 - \Phi(z_\alpha - \gamma/\sqrt{V})$$

where $\Phi(\cdot)$ *is the* $N(0,1)$ *CDF,* z_α *is the one side critical value of* $N(0,1)$ *at significance level* α*, and*

$$\begin{aligned} V &= 2\int \left[\iint |\Phi_y(u_1+u_2,x)|^2 dW_1(u_1)dW_1(u_2) \iint |\Phi_z(v_1+v_2,x)|^2 dW_2(v_1)dW_2(v_2)\right] a^2(x)dx \\ &\quad \times \int \left[\int K(\tau)K(\tau+\eta)d\tau\right]^2 d\eta \end{aligned} \tag{21}$$

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

Theorem 2 shows that our test has nontrivial power against the class of local alternatives $\mathbb{H}_1(a_n)$ with $a_n = n^{-1/2}h^{-d_x/4}$. We note that the convergence rate $n^{-1/2}h^{-d_x/4}$ is slower than the parametric rate $n^{-1/2}$, but only slightly. For example, if $h \propto n^{-\frac{1}{3+d_x}}$, then $n^{-1/2}h^{-d_x/4} = n^{-(6+d_x)/[4(3+d_x)]}$ is only slightly slower than $n^{-1/2}$. In contrast to Su and White (2007, 2008, 2011a), Bouezmarni et al. (2012) and many other smoothed nonparametric tests, the convergence rate of our test depends on the dimension of X_t only. As a result, our test is less severely subjected to the notorious "curse of dimensionality" problem and is asymptotically more efficient than the aforementioned tests. This is because our test only involves d_x dimensional smoothing, whereas the other tests involve d_x+d_z or $d_x+d_y+d_z$ dimensional smoothing. Thus, those tests can only detect local alternatives with a convergence rate of $n^{-1/2}h^{-(d_x+d_z)/4}$ or $n^{-1/2}h^{-(d_x+d_y+d_z)/4}$, which is slower than our rate of $n^{-1/2}h^{-d_x/4}$ and so is asymptotic less efficient than our test. In addition, our test is asymptotic pivotal under the null hypothesis for weakly dependent data. It is in stark contrast to Su and White's (2011b) test, which is only asymptotic pivotal for independent or martingale difference sequence data.

It should be noted that Su and White's (2011b) test can detect a class of local alternatives that converge to $\mathbb{H}_0$ at parametric rate $n^{-1/2}$, which is faster than $a_n = n^{-1/2}h^{-d_x/4}$ for our test. However, this conclusion is peculiar to the smooth type local alternatives in Eq.(20). Suppose we consider another class of local alternatives:

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

where $\delta(u,v,x)$ is a continuous function of x on G, c is a constant in the interior of the support of X_t, $a_n \to 0$, $b_n \to 0$ as $n \to \infty$, $a_n^2 b_n = n^{-1}h^{-d_x/2}$, and $h = o(b_n)$. This type of alternatives has been considered by Rosenblatt (1975) and Horowitz and Spokoiny (2001) among others.

The alternative $\mathbb{H}_2(a_n,b_n)$ could be transformed from some interesting local alternatives in terms of conditional density functions. For example, it can arise when

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

with

$$q(y,z|x) = q_a(y,z)G\left(\frac{x-c}{b_n}\right)$$

where $G(\cdot)$ is a bounded smooth function, $q_a(y,z) \neq 0$, and $\iint q_a(y,z)dydz = 0$. Under this kind of alternatives, the deviation between $\mathbb{H}_2(a_n, b_n)$ and $\mathbb{H}_0$ has a nonsmooth spike at location c. In this case, Y and Z display significant dependence conditional on X in a neighborhood of the value c. The shrinking 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 $\mathbb{H}_0$ for all x on its support 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 the order a_n for $X_t = c$, but of a higher order $a_n b_n$ for any other distinct points of X_t on its support except c.

Following an analogous proof of Theorem 2, we could obtain the asymptotic property of our test under the class of local alternatives $\mathbb{H}_2(a_n, b_n)$.

Theorem 3. *Suppose Assumptions A.1-A.4 hold and the bandwidth $h = cn^{-\lambda}$ for $1 \le d_x \le 8$ and $\frac{1}{d_x+4} < \lambda < \frac{2}{3d_x}$, 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 $h = o(b_n)$,*

$$P\left[\widehat{SM} \ge z_\alpha | \mathbb{H}_2(a_n, b_n)\right] \to 1 - \Phi(z_\alpha - \kappa/\sqrt{V})$$

where $\kappa = a(c)g(c)\iiint |\delta(u,v,x)|^2 dW_1(u)dW_2(v)dx$, $\Phi(\cdot)$ is the $N(0,1)$ CDF, z_α is the one side critical value of $N(0,1)$ at significance level α, and V is defined by Eq. (21).

With Theorem 3, we can show that our proposed test is asymptotically more efficient than not only Su and White (2007) but also Su and White (2011b) under $\mathbb{H}_2(a_n, b_n)$. For example, suppose $d_x = 1$, $h = n^{-\frac{1}{4}}$, $b_n = h^{\frac{5}{6}}$ and $a_n = n^{-1/2}h^{-2/3}$, the magnitude of the indefinite integral of $a_n\delta\left(u, v, (x-c)/b_n\right)$ over $x \in G$ is of the order $a_n b_n = n^{-1/2}h^{1/6}$, which vanishes to 0 faster than $n^{-1/2}$ given $h \to 0$. Thus, Su and White's (2011b) test will fail to detect the class of local alternatives $\mathbb{H}_2(a_n, b_n)$.

6 Inference on Patterns of Conditional Dependence

When the null hypothesis of conditional independence is rejected, one may like to gauge possible sources of rejection, which can provide valuable information for modeling the relationship among economic variables. For example, if we know that two variables have conditional dependence in mean, then we can search for a conditional mean model to capture it. Therefore, once the conditional independence hypothesis is rejected, inference on patterns of conditional dependence then becomes an important issue.

As is well known, the characteristic function can be differentiated to obtain various moments (if exists), which is quite useful in checking the existence of conditional dependence in various moments. As our ominous test is based on the conditional characteristic function, we will develop a class of derivative tests to capture various aspects of conditional dependence patterns. As we show below, our derivative tests are rather convenient to test various hypothesis of interest, including omitted variables, Granger causality in mean, and conditional uncorrelatedness. An important feature of our derivative tests is that they are all model-free, i.e., they do not impose auxiliary parametric restrictions when testing the hypotheses of interest. For example, we do not assume a parametric regression model when testing omitted variables.

6.1 Inference on Conditional Dependence of Various Moments

Suppose the p-th order moment of Y_t exists. For the generalized covariance function $\sigma(u,v,x)$ in Eq. (4), taking the p-th order partial derivative with respect to u at $u=0$, we obtain

$$\sigma^{(p)}(0,v,x) = \frac{\partial^p \sigma(u,v,x)}{\partial u^p}|_{u=0} = \mathbf{i}^p \mathrm{cov}\left(Y_t^p, e^{\mathbf{i}v'Z_t}|X_t = x\right), \tag{22}$$

for any $p = 1, 2, \cdots$.

Under the null hypothesis:

$$\mathbb{H}_0^{(p)} : P\left[\mathrm{cov}\left(Y_t^p, e^{\mathbf{i}v'Z_t}|X_t = x\right) = 0\right] = 1, \tag{23}$$

Eq. (22) equals to zero for all $x \in G$. Following Bierens (1982), we have $\mathrm{cov}(Y_t^p, e^{\mathbf{i}v'Z_t}|X_t = x) = 0$ if and only if $E(Y_t^p|X_t, Z_t) = E(Y_t^p|X_t)$. That is, we check whether Z_t could provide valuable information in modeling the mean dynamics of Y_t^p conditional on X_t. Thus, we can test (23) by examining the derivation of $\mathrm{cov}(Y_t^p, e^{\mathbf{i}v'Z_t}|X_t = x)$ from a zero function of x. Denote the nonparametric regression estimator of $\sigma^{(p)}(0,v,x)$ as $\hat{\sigma}^{(p)}(0,v,x) = \frac{\partial^p}{\partial u^p}\hat{\sigma}(u,v,x)|_{u=0}$. Similar to the construction of our ominous test, we can use the following quadratic form to test the null hypothesis (23):

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

Following the proof of Theorem 1, we can show that, under the null hypothesis $\mathbb{H}_0^{(p)}$ and other regularity conditions, the standardized version of $\hat{M}^{(p)}$ asymptotically follows a $N(0,1)$ distribution, i.e.,

$$\widehat{SM}^{(p)} = \frac{nh^{d_x/2}\hat{M}^{(p)} - \hat{B}^{(p)}}{\sqrt{\hat{V}^{(p)}}} \xrightarrow{d} N(0,1),$$

where

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

with $\hat{\phi}_y^{(s)}(0,x) = \frac{\partial^s}{\partial u^s}\hat{\phi}_y(u,x)|_{u=0}$, and $\hat{\Phi}_z(v,x)$ is defined as in Theorem 1.

Moreover, to improve size of the test statistic in finite samples, we can consider the following finite version test statistic:

$$\widehat{SM}_n^{(p)} = \frac{nh^{d_x/2}\hat{M}^{(p)} - \hat{B}_n^{(p)}}{\sqrt{\hat{V}^{(p)}}},$$

where

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

We now consider the primary case of $p = 1$. In this case, $\hat{M}^{(1)}$ will test the null hypothesis that $\text{cov}\left(Y_t, e^{iv'Z_t}|X_t\right) = 0$, which is equivalent to the model-free hypothesis $E(Y_t|X_t, Z_t) = E(Y_t|X_t)$, i.e., Z_t is not an omitted variable. Aït-Sahalia et al. (2001) also consider a nonparametric test for omitted variables in a time series regression context. In a cross-section context, Fan and Li (1996) and Lavergne and Vuong (2000) also develop some nonparametric tests for omitted variables using a weighted average of squared conditional mean estimates of residuals based on nonparametric smoothing. As our characteristic function based test only involves d_x dimensional smoothing, it is more powerful than tests of Aït-Sahalia et al. (2001), and Fan and Li (1996), Lavergne and Vuong (2000). The later involve a higher dimensional smoothing and are therefore asymptotically less efficient. In a time series context, our test $\hat{M}^{(1)}$ could be applied to test Granger causality in mean without any modification. Put $X_t = Y_{t-p}^{t-1} = (Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p})'$ and $Z_t = X_{t-q}^{t-1} = (X_{t-1}, X_{t-2}, \cdots, X_{t-q})'$. Then the null hypothesis becomes to $\mathbb{H}_0^{(1)} : E(Y_t|Y_{t-p}^{t-1}, X_{t-q}^{t-1}) = E(Y_t|Y_{t-p}^{t-1})$, i.e., there is no Granger causality in the mean of Y_t from X_{t-q}^{t-1}. Compared with the traditional F test for Granger causality, which assumes a linear regression model, our $\hat{M}^{(1)}$ test is a model-free test for Granger causality in mean, and it is powerful in capturing not only linear but also various nonlinear relationships in mean, including ARCH-in-mean effect (Engle et al. 1987), Threshold effect (Tong and Lim, 1980) and functional coefficient autoregressive model (Priestley, 1988; Chen and Tsay, 1993).

6.2 Inference on Conditional Correlation Between Two Specified Moments

Suppose the p-th and q-th order moments of Y_t and Z_t exist respectively. Then taking the p-th and q-th orders partial derivative of $\sigma(u, v, x)$ with respect to (u, v) at $(u, v) = (0, 0)$, we obtain

$$\sigma^{(p,q)}(0, 0, x) = \frac{\partial^{p+q}\sigma(u, v, x)}{\partial u^p \partial v^q}|_{(u,v)=(0,0)} = \mathbf{i}^{p+q}\text{cov}\,(Y_t^p, Z_t^q|X_t = x) \tag{24}$$

for any $p = 1, 2, \cdots ; q = 1, 2, \cdots$.

Under the null hypothesis:

$$\mathbb{H}_0^{(p,q)} : P\left[\text{cov}\,(Y_t^p, Z_t^q|X_t = x) = 0\right] = 1, \tag{25}$$

Eq. (24) equals to zero. Like in Section 6.1, we denote the nonparametric regression estimator of $\sigma^{(p,q)}(0, 0, x)$ as $\hat{\sigma}^{(p,q)}(0, 0, x) = \frac{\partial^{p+q}}{\partial u^p \partial v^q}\hat{\sigma}(u, v, x)|_{(u,v)=(0,0)}$. Then we could use the following statistic

$$\hat{M}^{(p,q)} = \frac{1}{n}\sum_{t=1}^{n} a(X_t)|\hat{\sigma}^{(p,q)}(0, 0, X_t)|^2 \tag{26}$$

to check conditional uncorrelatedness between Y_t^p and Z_t^q given X_t. Following an analogous reasoning to the proof of Theorem 1, we can prove that under the null hypothesis $\mathbb{H}_0^{(p,q)}$ and suitable regularity conditions, the standardized version of $\hat{M}^{(p,q)}$ also converges to a $N(0, 1)$ distribution, i.e.,

$$\widehat{SM}^{(p,q)} = \frac{nh^{d_x/2}\hat{M}^{(p,q)} - \hat{B}^{(p,q)}}{\sqrt{\hat{V}^{(p,q)}}} \xrightarrow{d} N(0, 1)$$

where

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

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

Once again, to improve the size of the test in finite samples, we can use a finite version of $\widehat{SM}^{(p,q)}$:

$$\widehat{SM}_n^{(p,q)} = \frac{nh^{d_x/2}\hat{M}^{(p,q)} - \hat{B}_n^{(p,q)}}{\sqrt{\hat{V}_n^{(p,q)}}}$$

where

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

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

with $\hat{e}_y^{(p)}(0, X_s) = Y_s^p - \hat{\phi}_y^{(p)}(0, X_s)$ and $e_z^{(q)}(0, X_s) = Z_s^q - \hat{\phi}_z^{(q)}(0, X_s)$.

The choice of derivative orders (p, q) allows us to examine various conditional correlation structures between Y_t and Z_t. As a primary example, we now consider the case of $(p, q) = (1, 1)$. This yields a model-free test of conditional uncorrelatedness with the null hypothesis $E(Y_t Z_t|X_t) = E(Y_t|X_t)E(Z_t|X_t)$. Su and Ullah (2009) also propose a nonparametric test of conditional uncorrelatedness in a time series context. Their test relies on the assumption of $E(Y_t|X_t) = E(Z_t|X_t) = 0$. Once this assumption fails, they first regress Y_t on X_t and Z_t on X_t nonparametrically and then construct the test statistic using the nonparametric residuals. Moreover, they carefully choose different bandwidths to avoid the impact of estimation errors from the first step on the asymptotic distribution of their test statistic. Compared with Su and Ullah (2009), we use a single bandwidth in estimating the conditional expectations $E(Y_t Z_t|X_t)$, $E(Y_t|X_t)$ and $E(Z_t|X_t)$, and allow the estimation errors from these conditional expectations to jointly determine the limiting distribution of our test. As a result, our test is expected to have a better size performance in finite samples because we have better asymptotic approximation.

For a time series $\{Y_t\}_{t=1}^n$ and a positive integer $2 \le k < n$, define $Z_t \triangleq Y_{t-k}$ and $X_t \triangleq Y_{t-k+1}^{t-1} = \{Y_{t-1}, \cdots, Y_{t-k+1}\}$, then $\gamma(k, y_{t-k+1}^{t-1}) \triangleq \text{cov}(Y_t, Y_{t-k}|Y_{t-k+1}^{t-1} = y_{t-k+1}^{t-1})$, where k is a lag order, is the well known partial autocovariance function (PACF) in time series analysis. Therefore, our test statistic $\hat{M}^{(1,1)}$ is a weighted average of squared PACFs and could be used to test the significance of the higher lag order of the dependent variable conditional on lower lag orders in a nonparametric autoregressive process. Compared with the commonly used t statistic, our nonparametric based test not only avoids misspecification problem, but also is powerful in detecting such nonlinear relationships as threshold and smooth transition autoregressive processes.

7 Monte Carlo Study

We now study the finite sample performance of the proposed tests in comparison with some popular tests of conditional independence, namely Su and White's (2007) test and Granger's (1969) F test for Granger causality. For the derivative tests $\hat{M}^{(p)}$ and $\hat{M}^{(p,q)}$, we consider the primitive cases of $p = 1$ and $(p, q) = (1, 1)$. As noted in Section 6, $\hat{M}^{(1)}$ tests whether Z_t is an omitted variable in modeling the conditional mean of Y_t given X_t, whereas $\hat{M}^{(1,1)}$ tests conditional uncorrelatedness between Y_t and Z_t given X_t.

To examine the size and power of $\hat{M}, \hat{M}^{(1)}, \hat{M}^{(1,1)}$ in finite samples, we consider the following DGPs:

$$
\begin{aligned}
\text{DGP.S1: } Y_t &= 0.5Y_{t-1} + \varepsilon_{1,t} \\
\text{DGP.S2: } Y_t &= \sqrt{h_t}\varepsilon_{1,t}, h_t = 0.01 + 0.5Y_{t-1}^2 \\
\text{DGP.S3: } Y_t &= \sqrt{h_{1,t}}\varepsilon_{1,t}, h_{1,t} = 0.01 + 0.9h_{1,t-1} + 0.05Y_{t-1}^2 \\
Z_t &= \sqrt{h_{2,t}}\varepsilon_{2,t}, h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2 \\
\text{DGP.P1: } Y_t &= 0.5Y_{t-1} + 0.5Z_{t-1} + \varepsilon_{1,t} \\
\text{DGP.P2: } Y_t &= 0.5Y_{t-1}Z_{t-1} + \varepsilon_{1,t} \\
\text{DGP.P3: } Y_t &= 0.4Y_{t-1} + 0.2Z_{t-1}^2 + \varepsilon_{1,t} \\
\text{DGP.P4: } Y_t &= 0.3 + 0.2\log(h_t) + \sqrt{h_t}\varepsilon_{1,t}, h_t = 0.01 + 0.5Y_{t-1}^2 + 0.3Z_{t-1}^2 \\
\text{DGP.P5: } Y_t &= 0.5Y_{t-1} + 0.5Z_{t-1}\varepsilon_{1,t} \\
\text{DGP.P6: } Y_t &= \sqrt{h_t}\varepsilon_{1,t}, h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Z_{t-1}^2 \\
\text{DGP.P7: } Y_t &= \sqrt{h_{1,t}}\varepsilon_{1,t}, h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4Y_{t-1}^2 + 0.5Z_{t-1}^2 \\
Z_t &= \sqrt{h_{2,t}}\varepsilon_{2,t}, h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2
\end{aligned}
$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are i.i.d.$N(0, 1)$ sequences and Z_t in DGP.S1-DGP.S2 and DGP.P1-DGP.P6 is generated by an AR(1) process

$$Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}.$$

All the above DGPs except DGP.P4 are investigated by Su and White (2008). DGP.P4 is an ARCH-in-mean process proposed by Engle et al. (1987). These DGPs cover a wide range of linear and nonlinear time series processes. In this section, we test whether Y_t is independent with Z_{t-1} conditional on Y_{t-1}, that is, whether Z_t Granger-causes Y_t by setting the lag order to 1. Among all ten DGPs, DGP.S1-S3 are used to study the sizes of our tests while DGP.P1-P7 allow us to examine their powers. All DGPs except DGP.P1 are nonlinear in mean or in variance or both. Under DGP. P3-P7, the null hypothesis of conditional uncorrelatedness for the test $\hat{M}^{(1,1)}$ holds and under DGP.P5-P7, the null hypothesis of no Granger causality in mean for the test $\hat{M}^{(1)}$ holds.

For each DGP, we simulate 1000 data sets with the sample size $n = 100, 200, 500, 1000$ respectively. For our tests $\hat{M}, \hat{M}^{(1)}, \hat{M}^{(1,1)}$, following Aït-Sahalia et al. (2001) and Chen and Hong (2010), we choose the Gaussian kernel $k(x) = 1/\sqrt{2\pi}\exp(-x^2/2)$ and the truncated weighting function $a(X_t) = \mathbf{1}(|X_t| \leq 1.5)$, where $\mathbf{1}(\cdot)$ is the indicator function and X_t has been standardized by its sample mean and standard deviation. We choose both $W_1(\cdot)$ and $W_2(\cdot)$ to be the $N(0,1)$ CDF and choose the bandwidth $h = n^{-2/9}$. We also consider the empirical sizes and powers of the tests under DGP.S1-S3 and DGP.P1-P3 respectively by setting $h = cn^{-2/9}$ with $c = 0.5, 1.5, 2$. The results are similar to those reported in Tables 1 and 2. For space, we do not report results with $c \neq 1$, which are available from the authors upon request.

For Su and White's (2007) test, we choose the fourth order kernel $k(u) = (2-u^2)\varphi(u)/2$, where $\varphi(u)$ is the $N(0,1)$ density. To make Su and White's (2007) test and ours comparable, we choose $h_1 = n^{-2/9}$ and $h_2 = n^{-1/3}$ for Su and White's (2007) test, which satisfy Assumption $A.2$ in Su and White (2007). We also consider Granger's (1969) F test for linear Granger causality in mean. For $n = 100, 200$, we use the local bootstrap procedure proposed by Paparoditis and Politis (2000) and modified by Su and White (2008). For the bootstrap, we generate 500 data sets for each DGP and use $B = 100$ bootstrap iterations for each simulated data set. In addition, we use the Gaussian kernel and h_1 as the bootstrap kernel and resampling bandwidth respectively.

Table 1 reports the sizes of tests under DGP.S1-S3 at the 10% and 5% significance levels using asymptotic critical values and bootstrap critical values respectively. All our three tests have reasonable sizes using both asymptotic critical values and bootstrap critical values. With asymptotic critical values, our tests tend to overreject a bit but not excessively, and they improve as the sample size n increases. This conforms the advantage of allowing the nonparametric estimation errors of both conditional joint and marginal characteristic functions to jointly determine the asymptotic distribution of the test statistic. The bootstrap procedure reduces overrejection, but the improvement is not significant. Since our tests have achieved reasonable sizes using asymptotic approximation, it should not be surprised to see the inappreciable role of bootstrap approximation. Thus, it does not seem to be necessary to use bootstrap for our tests. This is practically appealing, because bootstrap in a nonparametric time series context is rather time consuming. In contrast, Su and White's (2007) test suffers from severe overrejection when using asymptotic critical values but it has a remarkable improvement using the bootstrap procedure. Therefore, for Su and White's (2007) test, the bootstrap delivers more reliable results and we will use it in an empirical application below.

Table 2 reports the powers of Su and White's (2007) test, Granger's (1969) F test and our three tests under DGP.P1-P7 at the 10% and 5% levels, using asymptotic critical values and bootstrap critical values respectively. For simplicity, we only report the bootstrap results for Su and White's (2007) test and our test $\hat{M}$ when the sample size $n = 100, 200$ respectively. From Table 2, we see that the traditional F test is most powerful under DGP.P1, which has a linear Granger causality relationship. Our test $\hat{M}$ is very powerful in detecting all derivations given by DGP.P1-P7 and achieves unity power quickly as

n increases. For $\hat{M}$, the bootstrap power is slightly lower than the power based on asymptotic critical values, while for Su and White's (2007) test, the bootstrap power is significantly lower. In comparison with Su and White's (2007) test, $\hat{M}$ is generally more powerful in terms of both the asymptotic and bootstrap critical values. This is consistent with our analysis on the relative efficiency between our test and Su and White's (2007) test. Moreover, it is interesting to see that $\hat{M}^{(1)}$ is powerful in capturing various form of Granger causality in mean in DGP.P1-P4 and it is robust to higher order conditional dependence such as ARCH/GARCH effects in DGP.P5-P7 for which there exists no Granger causality in mean. Similarly, $\hat{M}^{(1,1)}$ is powerful in capturing various forms of conditional correlation between Y_t and Z_{t-1}, and is robust to conditional correlation in higher order moments. Indeed, under DGP.P3-P7, for which there exists no conditional correlation but there exists dependence in higher order moments, the empirical rejection frequencies of $\hat{M}^{(1,1)}$ are close to the nominal significance levels, which means that $\hat{M}^{(1,1)}$ has robust reasonable sizes under the null in finite samples. Finally, we note that $\hat{M}^{(1,1)}$ is different from the F test. The $\hat{M}^{(1,1)}$ test is powerful in capturing some nonlinear Granger relationships in mean such as DGP.P2, while the F test is silent about this kind of derivation.

8 Application to Nonlinear Granger Causality Between Money and Output

The relationship between money and output has attracted a phenomenal amount of interest over years from both empirical and theoretical macroeconomic studies. This issue not only reflects the causal relationship between nominal economic variables (such as money) and real economic variables (such as output), but also involves the discussion about whether the monetary policy is neutral. Since the 1970s, many studies have investigated the relationship between output and money, such as Sims (1972, 1980), Christiano and Ljungqvist (1988), Stock and Watson (1989), and Friedman and Kuttner (1993). However, the results vary with different sample intervals. Recently, some researchers believe that there exists a nonlinear relationship between the money and output, while the sensitivity testing result is a reflection of this nonlinear relationship. Indeed, a stream of economic theories imply a nonlinear relationship between money and output. The sources of nonlinear effect between money and output may include the nonlinear wage indexation and price adjustment (Kandil, 1995), the asymmetric preference of central bank's monetary policy (Nobay and Peel, 2003), the nonlinearity of aggregate supply and demand curve in economic reaction and so on. However, most related empirical studies have employed the traditional linear Granger causality test, which has little power in discovering nonlinear relationships, as seen in our simulation study. In this section, we will use our tests to study various Granger causality relationships between money and output.

We use US monthly data in the period 1959:M1-2012:M6, with 642 observations. We measure output by monthly Industrial Production Index (IPI). According to Psaradarkis et al. (2005), we use three monetary or financial variables, the narrow money supply M_1, the broad money supply M_2 and the Federal Funds rate (ir), as the proxy variables of monetary policy. We logarithmically transform $IPI,\ M_1,\ M_2$, denoted as $ipi,\ m_1,\ m_2$ respectively. All data except the interest rate are seasonally adjusted.

We first check the stationarity of the data by the augmented Dickey-Fuller test. The results suggest that $ipi,\ m_1,\ m_2,\ ir$ are integrated of order one, and the differenced series, which we denote as $\Delta ipi,\ \Delta m_1,\ \Delta m_2,\ \Delta ir$ are integrated of order zero. As mentioned in Bae and de Jone (2007), considering the fact that the Federal Reserve Broad usually adjusts its target interest rate by multiples 25 basis points, not by a certain percentage of the current interest level, it is more appropriate to assume its difference rather than the difference of its logarithm to be a stationary process. Thus, we can employ the Granger causality tests on the differenced series $\Delta ipi,\ \Delta m_1,\ \Delta m_2,\ \Delta ir$. The data series $\Delta ipi,\ \Delta m_1,\ \Delta m_2,\ \Delta ir$ are depicted in Figure 1.

The traditional linear Granger causality F test checks whether output (Δipi here) and money (Δm_1, Δm_2, Δir) Granger cause each other in the following linear regressions:

$$\Delta ipi_t = \alpha_0 + \alpha_1 \Delta ipi_{t-1} + \cdots + \alpha_p \Delta ipi_{t-p} + \beta_1 \Delta m_{t-1} + \cdots + \beta_q \Delta m_{t-q} + \varepsilon_{1t} \quad (27)$$

$$\Delta m_t = \alpha_0 + \alpha_1 \Delta m_{t-1} + \cdots + \alpha_p \Delta m_{t-p} + \beta_1 \Delta ipi_{t-1} + \cdots + \beta_q \Delta ipi_{t-q} + \varepsilon_{2t} \quad (28)$$

where Δm equals $\Delta m_1, \Delta m_2$ or Δir, and $\varepsilon_{it} \sim i.i.d.N(0, \sigma_i^2)$. The hypothesis of no Granger causality in mean for linear regressions (27) and (28) is :

$$\mathbb{H}_0 : \quad \beta_1 = \beta_2 = \cdots = \beta_q = 0.$$

Compared with the linear Granger causality F test, the null hypothesis of our $\hat{M}^{(1)}$ test is no Granger causality in mean:

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

and the null hypothesis of our $\hat{M}$ test is no Granger causality in distribution:

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

where $\Delta m_{t-s}^{t-1} = (\Delta m_{t-1}, \Delta m_{t-2}, \cdots, \Delta m_{t-s})$, and $\Delta ipi_{t-s}^{t-1} = (\Delta ipi_{t-1}, \Delta ipi_{t-2}, \cdots, \Delta ipi_{t-s})$, with $s = p, q$. The $\hat{M}^{(1)}$ test checks whether past money growths can provide valuable information in predicting the mean of future output growths and whether past output growths are useful in predicting the mean of future money growths. As documented in our simulation study, $\hat{M}^{(1)}$ is powerful in capturing linear and various nonlinear Granger causalities in mean, whereas the traditional F test is only powerful in detecting linear Granger causality. On the other hand, our $\hat{M}$ test checks whether past money growths are useful in predicting the distribution of future output growths and wether past output growths are useful in predicting the distribution of future money growths. Density forecasts for macroeconomic variables have been important for such decision makers as central banks (Diebold et al. 1999; Clements, 2004; Casillas-Olvera and Bessler, 2006).

We apply $\hat{M}$, Su and White's (2007) test, $\hat{M}^{(1)}$ and Granger's (1969) F test to investigate the Granger causalities between output and three monetary variables. All the data have been standardized to have zero mean and unit variance before applying these tests. For test statistics $\hat{M}$ and $\hat{M}^{(1)}$, we use the Gaussian kernel, the truncated weighting function $a(X_t) = \mathbf{1}(|X_t| \leq 1.5)$ and the $N(0,1)$ CDF for $W_i(.), i = 1, 2$. For the bandwidth, we set

$$h = h^* n^{-3/[2(4+d_x)]}$$

where d_x is the dimension of X_t, and h^* is the least-squares cross-validated bandwidth for estimating the conditional expectation of Y_t given X_t. For Su and White's (2007) test, we use a fourth order kernel as Su and White (2007). Since Su and White's (2007) test involves the choice of two bandwidths, we set

$$h_1 = h_1^* n^{-3/[2(4+d_x)]}, \qquad h_2 = h_2^* n^{-3/[4(d_x+4)]} n^{-3(d_x+d_z)/[4d_x(d_x+4)]}$$

where h_1^* and h_2^* are the least-squares cross-validated bandwidths for estimating the conditional expectation of Y_t given (X_t, Z_t) and X_t, respectively. These two bandwidths satisfy Assumption $A.2$ in Su and White's (2007). As the bootstrap procedure is more reliable than the asymptotic distribution for Su and White's (2007) test, we use the bootstrap in this empirical application. The Gaussian kernel is used as the bootstrap kernel and the resampling bandwidth is:

$$h_b = n^{-1/(d_x(d_x+4))}$$

The resampling bandwidth satisfies Assumption $A.8$ in Paparoditis and Politis (2000) and $h_b = n^{-1/5}$, when $d_x = 1$. We use $B = 200$ bootstrap iterations and choose the least-squares cross-validated bandwidths for each iteration. We also obtain results using the following two bandwidths: (1) We fixed $h^* = h_1^* = h_2^* = d_x$ for both the original data series and bootstrap samples; (2) We select the least-squares cross-validated bandwidths using the original data and regarded them as the fixed bandwidths for bootstrap samples. The bootstrap iterations $B = 500$ for these two cases. The results are rather similar to Table 3 and is available by request to the author. Besides, for Su and White's (2007) test, we further obtain results using the bandwidths given by Su and White (2007) and the conclusion has no significant difference with which based on Panel C of Table 3.

The results of the tests are summarized in Table 3, where we choose the lag orders $p, q = 1, 2, 3$ respectively. Panel A of Table 3 reports the results of Granger's (1969) F test. At the 5% level, all of three monetary variables do not Granger cause output, which indicates the ineffectiveness of monetary policy. This result is consistent with Uhlig's (2005) linear VAR setup based conclusion that monetary policy shocks have no clear effect on real GDP. Besides, the results of Panel A also suggest that the growth rate of M2 does not respond to the growth rate of output, and the growth rate of M1 only respond to output at the third order lag. However, it rejects the null hypothesis that Δipi does not Granger cause Δir for any lags, which may indicate the existence and rationality of the linear Taylor rule (Taylor, 1993).

Panel B of Table 3 reports the results of the $\hat{M}^{(1)}$ test. Compared with Panel A, our $\hat{M}^{(1)}$ test reveals further Granger causalities in mean between money and output in addition to the traditional F test. For example, according to our $\hat{M}^{(1)}$ test, Δipi Granger causes Δm_1 at the second and third lag orders, and Granger causes Δm_2 at the third lag order. Thus, the results of Panel B document the existence of nonlinear Granger causalities in mean and provide justification for modeling the relationship of money and output by a nonlinear conditional mean model.

The results of Su and White's (2007) test and our ominous $\hat{M}$ test are given in Panels C and D of Table 3 respectively. Comparing Panel C with Panel A, we find no significant difference between Su and White's (2007) test and the traditional Granger (1969)'s F test. That is, Su and White's (2007) test can not detect any additional relationship between money and output except the linear Granger causality. However, from the left part of Panel D, our $\hat{M}$ test documents strong evidence against the null hypothesis that output does not Granger cause money for all of the three monetary variables and all lag orders except for Δm_2 when $p = 1$. This result implies that the monetary authority responds to economic situations and uses some appropriate monetary polices to stimulate recovery or curb overheating. In addition, the right part of Panel D shows that interest rate is effective in stimulating the economy for any lag orders $p, q = 1, 2, 3$, and there is one month lag for broad money supply to affect output. Besides, we do not find any evidence against ineffectiveness of narrow money supply in affecting the economy. With the development of the financial markets and the convenience of borrowing, it is not difficult to understand the ineffectiveness of the narrow money supply. To sum up, the results of our test indicate strong evidence against the non-Granger causality between money and output for most cases. This is consistent with the recent use of nonlinear models in capturing the relationship between money and output in the literature.

9 Conclusion

Conditional independence is one of most widely used concepts in economic and financial modeling and encompasses many important assumptions in econometrics and statistics, such as the Markov property, Granger causality, missing at random and exogeneity. In this paper, we propose a test for conditional independence via a nonparametric regression approach in combination with the use of conditional characteristic function. In comparison with Su and White's (2007) test and other nonparametric approaches in the literature, our test has the following appealing features: our test is asymptotically locally more powerful in detecting a class of local alternatives; it is more flexible in inferring possible patterns of conditional dependence and it does not require use of a higher order kernel and multi-bandwidths. By adopting a single bandwidth, we allow the nonparametric estimation errors of both conditional joint and marginal characteristic functions to jointly determine the asymptotic distribution of our test statistic. As a result, our test has much better size than Su and White (2007) in finite samples. In addition, by taking appropriate order partial derivatives, our test can be used to construct model-free tests for omitted variables, Granger causality in mean, and conditional uncorrelatedness. All of the proposed tests have a convenient null asymptotic one-side $N(0, 1)$ distribution.

Monte carlo simulation study shows that our tests have reasonable size and excellent power in comparison with Su and White's (2007) test and the traditional F test for Granger causality. More importantly, it does not seem to be necessary to use bootstrap for our test, which is practically appealing because nonparametric bootstrap is very time consuming. We apply our tests to study the Granger causality between money and output. The results of our tests document some nonlinear relationships which are ignored by the traditional Granger causality test and Su and White's (2007) test. They provide justification on necessity of modeling the relationship between money and output via nonlinear models.

References

Aït-Sahalia Y., Bickel P.J. and Stocker T.M., 2001, Goodness-of -fit Tests for Kernel Regression with an Application to Option Implied Volatilities, *Journal of Econometrics*, 105, 363-412.

Aït-Sahalia Y., Fan J. and Jiang J., 2010, Nonparametric Tests of the Markov Hypothesis in Continuous-time Models, *Annals of Statistics*, 38, 3129-3163.

Aït-Sahalia Y., Fan J. and Peng H., 2009, Nonparametric Transition-Based Tests for Jump Diffusions, *Journal of the American Statistical Association*, 104, 1102-1116.

Bae Y. and de Jone R., 2007, Money Demand Function Estimation by Nonlinear Cointegration, *Jornal of Applied Econometrics*, 22, 767-793.

Beja A., 1979, State Preference and the Riskless Interest Rate: A Markov Model of Capital Markets, *Review of Economic Studies*, 46(3), 435-446.

Bierens H.J., 1982, Consistent Model Specification Tests, *Journal of Econometrics*, 20, 105-134.

Blundell R. and Horowitz J.L., 2007, A Non-Parametric Test of Exogeneity, *Review of Economic Studies*, 74, 1035-1058.

Bouezmarni T., Rombouts J.V.K. and Taamouti A., 2012, Nonparametric Copula-Based Test for Conditional Independence with Applications to Granger Causality, *Journal of Business & Economic Statistics*, 30(2), 275-287.

Bouissou B. M., Laffont J.J. and Vuong Q.H., 1986, Test of Noncausality Under Markov Assumptions for Qualitative Panel Data, *Econometrica*, 54, 395-414.

Calderón C. and Liu L., 2003, The Direction of Causality between Financial Development and Economic Growth, *Journal of Development Economics*, 72, 321-334.

Campbell J.Y., 1992, No News Is Good News: An Asymmetric Model of Changing Volatility in Stock Returns, *Journal of Financial Economics*, 31(3), 281-318.

Casillas-Olvera G. and Bessler D.A., 2006, Probability Forecasting and Central Bank Accountability, *Journal of Policy Modeling*, 28, 223-234.

Chen B. and Hong Y., 2010, Characteristic Function-based Testing for Multifactor Continuous-time Markov Models via Nonparametric Regression, *Econometric Theory*, 26, 1115-1179.

Chen B. and Hong Y., 2012, Testing for the Markov Property in Time Series, *Econometric Theory*, 28, 130-178.

Chen R. and Tsay R.S., 1993, Functional-Coefficient Autoregressive Models, *Journal of the American Statistical Association*, 88, 298-308.

Christiano L.J. and Ljungqvist L., 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, *The Economic Journal*, 114, 844-866.

Dawid A.P., 1979, Conditional Independence in Statistical Theory, *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(1), 1-31.

Dawid A.P., 1980, Conditional Independnece for Statistical Operations, *The Annals of Statistics*, 8(3), 598-617.

Diebold F.X., Hahn J. and Tay A.S., 1999, Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-frequency Returns of Foreign Exchange, *Review of Economics and Statistics*, 81(4), 661-673.

Easley D. and O'Hara M., 1987, Price, Trade size, and Information in Securities Markets, *Journal of Financial Economics*, 19, 69-90.

Engle R.F., Lilien D.M. and Robins R.P., 1987, Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M model, *Econometrica*, 55(2), 391-407.

Fan J. and Zhang J., 2004, Sieve Empirical Likelihood Ratio Tests for Nonparametric Functions, *Annals of Statis-*

tics, 32(5), 1858-1907.

Fan Y. and Li Q., 1996, Consistent Model Specification Tests: Omitted Varibales and Semiparametric Function Forms, *Econometrica*, 64, 865-890.

Fan Y. and Li Q., 1999, Root$-n-$consistent Estimation of Partially Linear Time Series Models, *Journal of Nonparametric Statistics*, 10, 245-271.

Fridman B.M. and Kutter K.N., 1993, Another Look at the Evidence on Money-income Causality, *Journal of Econometrics*, 57, 189-203.

Granger C.W.J., 1969, Investigating Causal Relations by Econometric Models and Cross-spectral Methods, *Econometrica*, 37(3), 424-438.

Granger C.W.J., 1980, Testing for cCausality: a Personal Viewpoint, *Journal of Economic Dynamics & Control*, 2, 329-352.

Granger C.W.J., Huangb B.N. and Yang C.W., 2000, A Bivariate Causality between Stock Prices and Exchange Rates: Evidence from Recent Asian flu, *The Quarterly Review of Economics and Finance*, 40(3), 337-354.

Hahn J., Todd P. and Klaauw W.V., 2001, Identification and Estimation of Treatemnt Effects with a Regression-discontinuity Design, *Econometrica*, 69(1), 201-209.

Hall P. and Horowitz J.L., 2005, Nonparametric Methods for Inference in the Presence of Instrumental variables, *Annals of Statistics*, 33, 2904-2929.

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., Yao Q. and Tjøstheim D., 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 the American Statistical Association*, 94, 1201-1220.

Hong Y. and Lee Y., 2013, A Loss Function Approach to Model Specification Testing and Its Relative Efficiency, *Annals of Statistics*, Forthcoming.

Horowitz J.L. and Manski C.F., 2000, Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data, *Journal of the American Statistical Association*, 95, 77-84.

Horowitz J.L. and Spokoiny V.G., 2001, An Adaptive, Rate-optimal Test of a Parametric Mean-regression Model Against a Nonparametric Alternative, *Econometrica*, 69(3), 599-631.

Huang T.M., 2010, Testing Conditional Independence Using Maximal Nonlinear Conditional Correlation, *The Annals of Statistics*, 38(4), 2047-2091.

Kandil M., 1995, Asymmetric Norminal Flexibility and Economic Fluctuations, *Southern Economic Journal*, 61, 674-695.

Kim C.J. and Nelson C.R., 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. and Vuong Q., 2000, Nonparametric Siginificance Testing, *Econometric Theory*, 16, 576-601.

Lee J., 2009, A Consistent Nonparametric Bootstrap Test of Exogeneity,*Working Paper*.

Li Q. and Racine J.S., 2006, *Nonparametric Econometrics: Theory and Practice*, Princeton: Princeton University Press.

Linton O. and Gozalo P., 1996, Conditional Independence Restrictions: Testing and Estimation, *Cowles Foundation Discussion Paper NO. 1140*.

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*, New York: Springer-Verlag.

Manski C. F., 2007, *Identification for Prediction and Decision*, Princeton: Princeton University Press.

Nobay R.A. and Peel D.A., 2003, Optimal Discretionary Monetary Policy in a Model of Asymmetric Central Bank Preferences, *Economic Journal*, 113, 657-665.

Pakes A. and McGuire P., 2001, Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the 'Curse' of Dimensionality, *Econometrica*, 69, 1261-1281.

Paparodits E. and Politis D.N., 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 & Finance*, 23(6), 847-862.

Priestley M.B., 1988, *Nonlinear and Nonstationary Time Series Analysis*, New York: Academic Press.

Psaradakis Z., Ravn M. O. and Sola M., 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, *The 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, *Handbook of Econometrics*, 4, 3081-3143.

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.

Stock J.H. and Watson M.W., 1989, Interpretating the Evidence on Money-income Causality, *Journal of Econometrics*, 40, 161-181.

Su L. and Ullah A., 2009, Testing Conditional Uncorrelatedness, *Journal of Business & Economic Statistics*, 27(1), 18-29.

Su L. and White H., 2007, A Consistent Characteristic Function-based Test for Conditional Independence, *Journal of Econometrics*, 141, 807-834.

Su L. and White H., 2008, A Nonparametric Hellinger Metric Test for Conditional Independence, *Econometric Theory*, 24, 829-864.

Su L. and White H., 2011a, Testing Conditional Independence via Empirical Likelihood, *working paper*.

Su L. and White H., 2011b, Conditional Independence Specification Testing for Dependent Process with Local Polynomial Quantile Regression, *working paper*.

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.

Thornton D.L. and Batten D.S., 1985, Lag-Lenghth Selection and Tests of Granger Causality between Money and Income, *Journal of Money, Credit and Banking*, 17(2), 164-178.

Tong H. and Lim K.S., 1980, Threshold Autoregression, Limit Cycles and Cyclical Data, *Journal of the Royal Statistical Society*, 42(3), 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., Linton O. and Hardle W., 2004, Semiparametric Regression Analysis with Missing Response at Random, *Journal of the American Statistical Association*, 99, 334-345.

Table 1: Size of Tests Under DGP.S1-S3

		$SW07$		$\hat{M}$		$\hat{M}^{(1)}$		$\hat{M}^{(1,1)}$		LIN	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.S1	$n = 100, AS$	.247	.344	.093	.148	.097	.155	.108	.149	.050	.106
	$n = 100, BS$	.048	.096	.038	.092	.044	.092	.038	.088	–	–
	$n = 200, AS$	.198	.288	.073	.117	.070	.121	.060	.094	.052	.102
	$n = 200, BS$	.048	.086	.042	.086	.032	.090	.052	.110	–	–
	$n = 500, AS$	.156	.240	.056	.103	.074	.122	.069	.107	.045	.092
	$n = 1000, AS$	.137	.234	.072	.114	.069	.108	.070	.105	.046	.093
DGP.S2	$n = 100, AS$	.247	.354	.120	.177	.084	.140	.102	.144	.047	.103
	$n = 100, BS$	.068	.126	.068	.126	.060	.106	.054	.110	–	–
	$n = 200, AS$	.238	.337	.083	.135	.066	.107	.072	.114	.057	.108
	$n = 200, BS$	.054	.092	.048	.098	.056	.116	.078	.116	–	–
	$n = 500, AS$	.194	.292	.067	.112	.057	.087	.064	.100	.037	.100
	$n = 1000, AS$	.173	.217	.070	.109	.063	.103	.068	.115	.050	.086
DGP.S3	$n = 100, AS$	.224	.313	.107	.166	.076	.121	.072	.107	.042	.094
	$n = 100, BS$	.040	.072	.054	.090	.050	.092	.064	.124	–	–
	$n = 200, AS$	.199	.296	.072	.116	.061	.108	.064	.106	.046	.102
	$n = 200, BS$	.040	.080	.062	.090	.078	.130	.066	.114	–	–
	$n = 500, AS$	.167	.254	.064	.107	.071	.107	.062	.103	.053	.097
	$n = 1000, AS$	.123	.203	.072	.114	.063	.101	.067	.107	.051	.101

Notes: (i) Results of $SW07$ are based on Su and White's (2007) test, and results of LIN are Granger's (1969) F test for linear Granger causality. (ii) AS and BS denote the results using asymptotic critical values and bootstrap critical values respectively. (iii) The results using asymptotic critical values are based on 1000 iterations, while the bootstrap results are based on 500 iterations.

Table 2: Power of Tests Under DGP.P1-P7

		$SW07, AS$		$SW07, BS$		$\hat{M}, AS$		$\hat{M}, BS$		$\hat{M}^{(1)}, AS$		$\hat{M}^{(1,1)}, AS$		LIN	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.P1	$n = 100$	.840	.886	.436	.572	.978	.984	.914	.968	.990	.991	.997	1.00	1.00	1.00
	$n = 200$	.970	.985	.794	.876	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$n = 500$	1.00	1.00	–	–	1.00	1.00	–	–	1.00	1.00	1.00	1.00	1.00	1.00
	$n = 1000$	1.00	1.00	–	–	1.00	1.00	–	–	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P2	$n = 100$	.730	.806	.420	.526	.918	.953	.836	.896	.952	.965	.957	.973	.206	.290
	$n = 200$	.885	.917	.694	.778	.996	.997	.996	.996	.998	.999	.999	1.00	.215	.286
	$n = 500$	.997	.997	–	–	1.00	1.00	–	–	1.00	1.00	1.00	1.00	.249	.323
	$n = 1000$	1.00	1.00	–	–	1.00	1.00	–	–	1.00	1.00	1.00	1.00	.265	.338
DGP.P3	$n = 100$	.397	.492	.102	.170	.476	.588	.316	.438	.607	.677	.139	.189	.175	.267
	$n = 200$	.462	.583	.164	.264	.743	.815	.618	.716	.850	.903	.084	.134	.149	.219
	$n = 500$	.669	.772	–	–	.986	.991	–	–	.998	1.00	.082	.118	.165	.253
	$n = 1000$	.899	.945	–	–	1.00	1.00	–	–	1.00	1.00	.069	.100	.181	.252
DGP.P4	$n = 100$	.877	.925	.568	.688	.790	.878	.554	.712	.341	.443	.078	.133	.191	.264
	$n = 200$	.989	.994	.882	.936	.973	.985	.904	.956	.481	.561	.058	.106	.149	.227
	$n = 500$	1.00	1.00	–	–	1.00	1.00	–	–	.799	.867	.063	.100	.189	.264
	$n = 1000$	1.00	1.00	–	–	1.00	1.00	–	–	.987	.994	.056	.101	.173	.255
DGP.P5	$n = 100$	.967	.985	.760	.864	.967	.984	.884	.950	.208	.284	.086	.135	.250	.344
	$n = 200$	1.00	1.00	.982	.994	1.00	1.00	1.00	1.00	.197	.260	.066	.104	.223	.310
	$n = 500$	1.00	1.00	–	–	1.00	1.00	–	–	.189	.236	.059	.091	.248	.326
	$n = 1000$	1.00	1.00	–	–	1.00	1.00	–	–	.172	.216	.062	.093	.267	.340
DGP.P6	$n = 100$	.776	.858	.390	.524	.689	.801	.494	.650	.151	.223	.083	.138	.163	.243
	$n = 200$	.922	.959	.610	.736	.937	.969	.850	.932	.129	.190	.070	.104	.147	.227
	$n = 500$	1.00	1.00	–	–	1.00	1.00	–	–	.135	.182	.061	.092	.195	.280
	$n = 1000$	1.00	1.00	–	–	1.00	1.00	–	–	.096	.144	.061	.092	.178	.255
DGP.P7	$n = 100$	.664	.766	.302	.424	.538	.646	.354	.518	.140	.189	.081	.134	.175	.268
	$n = 200$	.787	.872	.436	.602	.800	.889	.690	.800	.123	.186	.066	.098	.215	.167
	$n = 500$	.979	.991	–	–	1.00	1.00	–	–	.108	.157	.053	.088	.167	.249
	$n = 1000$	.998	1.00	–	–	1.00	1.00	–	–	.097	.152	.051	.089	.157	.241

Notes: (i) Results of $SW07$ are based on Su and White (2007)'s test, and results of LIN are the traditional F test for linear model. (ii) AS and BS denote the results using asymptotic critical value and bootstrap critical value respectively. (iii) The results using asymptotic critical value are based on 1000 iterations, while the bootstrap results are based on 500 iterations.

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
Panel A: Granger's (1969) linear Granger causality F test						
$p = 1, q = 1$	.798	.557	.000	.332	.491	.034
$p = 1, q = 2$	.144	.232	.000	.581	.739	.052
$p = 1, q = 3$	.000	.098	.000	.684	.329	.111
$p = 2, q = 1$	.848	.466	.000	.364	.432	.094
$p = 2, q = 2$	.282	.232	.000	.662	.679	.188
$p = 2, q = 3$	.000	.089	.000	.722	.341	.296
$p = 3, q = 1$	.855	.413	.000	.280	.336	.139
$p = 3, q = 2$	.218	.185	.000	.551	.579	.302
$p = 3, q = 3$	.000	.071	.000	.714	.286	.349
Panel B: This paper's nonlinear Granger causality in mean test $\hat{M}^{(1)}$						
$p = 1, q = 1$	.165	.220	.055	.430	.625	.035
$p = 1, q = 2$	.000	.175	.010	.685	.295	.015
$p = 1, q = 3$	.000	.025	.020	.385	.145	.085
$p = 2, q = 1$	.085	.175	.005	.660	.660	.145
$p = 2, q = 2$	.000	.085	.000	.830	.520	.130
$p = 2, q = 3$	.000	.035	.000	.400	.340	.150
$p = 3, q = 1$	.105	.180	.010	.500	.365	.085
$p = 3, q = 2$	.015	.020	.005	.690	.230	.060
$p = 3, q = 3$	.000	.025	.010	.285	.160	.085
Panel C: Su and White's (2007) nonlinear Granger causality test						
$p = 1, q = 1$	.345	.185	.000	.885	.525	.000
$p = 1, q = 2$	.580	.865	.015	.890	.275	.720
$p = 1, q = 3$	.180	.995	.055	.970	.865	.845
$p = 2, q = 1$	.370	.230	.005	.070	.050	.005
$p = 2, q = 2$	.105	.530	.000	.645	.715	.855
$p = 2, q = 3$	.315	.850	.020	.530	.895	.710
$p = 3, q = 1$	.000	.075	.005	.750	.290	.080
$p = 3, q = 2$	.095	.170	.025	.105	.270	.190
$p = 3, q = 3$	.135	.480	.015	.030	.280	.565
Panel D: This paper's nonlinear Granger causality test $\hat{M}$						
$p = 1, q = 1$	.025	.075	.015	.425	.130	.000
$p = 1, q = 2$	.000	.075	.020	.450	.015	.000
$p = 1, q = 3$	.000	.135	.040	.150	.025	.010
$p = 2, q = 1$	.000	.045	.000	.415	.145	.000
$p = 2, q = 2$	.000	.040	.000	.310	.045	.000
$p = 2, q = 3$	.000	.025	.000	.150	.045	.005
$p = 3, q = 1$	.050	.050	.000	.520	.110	.000
$p = 3, q = 2$	.010	.005	.000	.595	.020	.000
$p = 3, q = 3$	.005	.015	.000	.105	.010	.000

Notes: (i) Numbers in the main entries are the $p-$values. (ii) $p-$values of the linear Granger causality test are calculated using F distribution with $(q, n-p-q)$ degrees of freedom. (iii) $p-$values of Su and White's (2007) test and our tests are based on 200 bootstrap iterations.

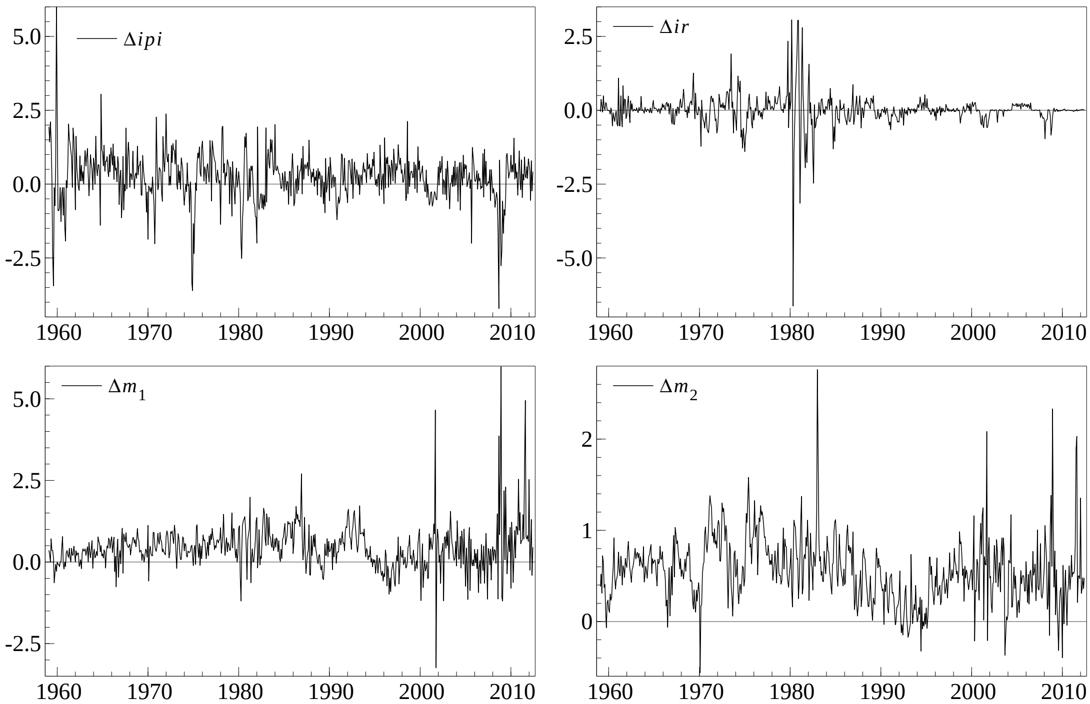

Figure 1: Data Series During the Period 1959:M01- 2012:M06

Appendix

Thoughout the appendix, we denote $d_x = k$, and

$$\begin{aligned}
\hat{M}_h = nh^{k/2}\hat{M} &= h^{k/2}\sum_{t=1}^{n}\iint |\hat{\sigma}(u,v,X_t)|^2 a(X_t)dW_1(u)dW_2(v),\\
\varepsilon_{yz}(u,v,X_s) &= e^{\mathbf{i}(u'Y_s+v'Z_s)} - \phi_{yz}(u,v,X_s),\\
\varepsilon_y(u,X_s) &= \varepsilon_{yz}(u,0,X_s),\\
\varepsilon_z(v,X_s) &= \varepsilon_{yz}(0,v,X_s).
\end{aligned}$$

In addition, $c \in (0,\infty)$ is a generic bounded constant that may vary from case to case. $A \sim B$ means A and B are the same order of magnitude.

Proof of Theorem 1. We first decompose $\hat{\sigma}(u,v,x)$ into four terms:

$$\begin{aligned}
\hat{\sigma}(u,v,x) &= \hat{\phi}_{yz}(u,v,x) - \hat{\phi}_y(u,x)\hat{\phi}_z(v,x)\\
&= \left[\hat{\phi}_{yz}(u,v,x) - \phi_{yz}(u,v,x)\right] - \phi_z(v,x)\left[\hat{\phi}_y(u,x) - \phi_y(u,x)\right]\\
&\quad -\phi_y(u,x)\left[\hat{\phi}_z(v,x) - \phi_z(v,x)\right] - \left[\hat{\phi}_y(u,x) - \phi_y(u,x)\right]\left[\hat{\phi}_z(v,x) - \phi_z(v,x)\right]. \qquad (29)
\end{aligned}$$

According to Eq. (29), $\hat{M}_h$ could be decomposed into ten terms:

$$\begin{aligned}
\hat{M}_h &= h^{k/2}\sum_{t=1}^{n}\iint\Big\{|\hat{\phi}_{yz}-\phi_{yz}|^2 + |\phi_y|^2|\hat{\phi}_z-\phi_z|^2 + |\phi_z|^2|\hat{\phi}_y-\phi_y|^2 + 2Re\left[\phi_y\phi_z^*(\hat{\phi}_z-\phi_z)(\hat{\phi}_y-\phi_y)^*\right]\\
&\quad -2Re\left[(\hat{\phi}_{yz}-\phi_{yz})\hat{\phi}_y^*(\hat{\phi}_z-\phi_z)^*\right] - 2Re\left[(\hat{\phi}_{yz}-\phi_{yz})\phi_z^*(\hat{\phi}_y-\phi_y)^*\right] + |(\hat{\phi}_y-\phi_y)(\hat{\phi}_z-\phi_z)|^2\\
&\quad -2Re\left[(\hat{\phi}_{yz}-\phi_{yz})(\hat{\phi}_y-\phi_y)^*(\hat{\phi}_z-\phi_z)^*\right] + 2Re\left[(\hat{\phi}_y-\phi_y)\phi_y^*\right]|\hat{\phi}_z-\phi_z|^2\\
&\quad +2Re\left[(\hat{\phi}_z-\phi_z)\phi_z^*\right]|\hat{\phi}_y-\phi_y|^2\Big\}a(X_t)dW_1(u)dW_2(v)\\
&= T_1+T_2+T_3+T_4+T_5+T_6+T_7+T_8+T_9+T_{10}, \qquad (30)
\end{aligned}$$

where $\hat{\phi}_{yz} \equiv \hat{\phi}_{yz}(u,v,x)$, $\hat{\phi}_y \equiv \hat{\phi}_y(u,x)$, $\hat{\phi}_z \equiv \hat{\phi}_z(v,x)$, $\phi_{yz} \equiv \phi_{yz}(u,v,x)$, $\phi_y \equiv \phi_y(u,x)$, $\phi_z \equiv \phi_z(v,x)$. We should analyze these ten terms T_1 to T_{10} given by Eq. (30) one by one to extract leading terms that determine the asymptotic distribution of our test. The leading terms of T_1 to T_{10} are given by Propositions 1 to 7 as follows.

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

$$T_1 = B_1 + \tilde{U}_1 + o_P(1),$$

with

$$\begin{aligned}
B_1 &= h^{-k/2}\iiint a(x)\left[1-|\phi_{yz}(u,v,x)|^2\right]dW_1(u)dW_2(v)dx\int K(\tau)^2 d\tau,\\
\tilde{U}_1 &= \frac{2}{nh^{3k/2}}\sum_{1\le s<r\le n} U_1(\xi_s,\xi_r)\\
&= \frac{2}{nh^{3k/2}}\sum_{1\le s<r\le n}\iiint \frac{a(x)}{g(x)}K\left(\frac{X_s-x}{h}\right)K\left(\frac{X_r-x}{h}\right)\\
&\quad \times Re\left[\varepsilon_{yz}(u,v,X_s)\varepsilon_{yz}(u,v,X_r)^*\right]dW_1(u)dW_2(v)dx.
\end{aligned}$$

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

$$T_2 = B_2 + \tilde{U}_2 + o_P(1),$$

with

$$\begin{aligned}
B_2 &= h^{-k/2} \iiint a(x)|\phi_y(u,x)|^2 \left[1 - |\phi_z(v,x)|^2\right] dW_1(u)dW_2(v)dx \int K(\tau)^2 d\tau, \\
\tilde{U}_2 &= \frac{2}{nh^{3k/2}} \sum_{1\le s<r\le n} U_2(\xi_s, \xi_r) \\
&= \frac{2}{nh^{3k/2}} \sum_{1\le s<r\le n} \iiint \frac{a(x)}{g(x)} K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right) |\phi_y(u,x)|^2 \\
&\quad \times Re\left[\varepsilon_z(v, X_s)\varepsilon_z(v, X_r)^*\right] dW_1(u)dW_2(v)dx.
\end{aligned}$$

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

$$T_3 = B_3 + \tilde{U}_3 + o_P(1),$$

with

$$\begin{aligned}
B_3 &= h^{-k/2} \iiint a(x)|\phi_z(v,x)|^2 \left[1 - |\phi_y(u,x)|^2\right] dW_1(u)dW_2(v)dx \int K(\tau)^2 d\tau, \\
\tilde{U}_3 &= \frac{2}{nh^{3k/2}} \sum_{1\le s<r\le n} U_3(\xi_s, \xi_r) \\
&= \frac{2}{nh^{3k/2}} \sum_{1\le s<r\le n} \iiint \frac{a(x)}{g(x)} K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right) |\phi_z(v,x)|^2 \\
&\quad \times Re\left[\varepsilon_y(u, X_s)\varepsilon_y(u, X_r)^*\right] dW_1(u)dW_2(v)dx.
\end{aligned}$$

Proposition 4. *Under the conditions of Theorem 1,*

$$T_4 = \tilde{U}_4 + o_P(1),$$

with

$$\begin{aligned}
\tilde{U}_4 &= \frac{2}{nh^{3k/2}} \sum_{s\ne r} U_4(\xi_s, \xi_r) \\
&= \frac{2}{nh^{3k/2}} \sum_{s\ne r} \iiint \frac{a(x)}{g(x)} K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right) \\
&\quad \times Re\left[\phi_y(u,x)\phi_z(v,x)^*\varepsilon_z(v, X_s)\varepsilon_y(u, X_r)^*\right] dW_1(u)dW_2(v)dx.
\end{aligned}$$

Proposition 5. *Under the conditions of Theorem 1,*

$$T_5 = B_5 + \tilde{U}_5 + o_P(1),$$

with

$$\begin{aligned}
B_5 &= -2h^{-k/2} \iiint a(x)|\phi_y(u,x)|^2 \left[1 - |\phi_z(v,x)|^2\right] dW_1(u)dW_2(v)dx \int K(\tau)^2 d\tau, \\
\tilde{U}_5 &= \frac{2}{nh^{3k/2}} \sum_{s\ne r} U_5(\xi_s, \xi_r) \\
&= -\frac{2}{nh^{3k/2}} \sum_{s\ne r} \iiint \frac{a(x)}{g(x)} K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right) \\
&\quad \times Re\left[\phi_y(u,x)^*\varepsilon_{yz}(u, v, X_s)\varepsilon_z(v, X_r)^*\right] dW_1(u)dW_2(v)dx.
\end{aligned}$$

Proposition 6. *Under the conditions of Theorem 1,*

$$T_6 = B_6 + \tilde{U}_6 + o_P(1),$$

with

$$\begin{aligned}
B_6 &= -2h^{-k/2} \iiint a(x)|\phi_z(v,x)|^2 \left[1 - |\phi_y(u,x)|^2\right] dW_1(u)dW_2(v)dx \int K(\tau)^2 d\tau, \\
\tilde{U}_6 &= \frac{2}{nh^{3k/2}} \sum_{s \neq r} U_6(\xi_s, \xi_r) \\
&= -\frac{2}{nh^{3k/2}} \sum_{s \neq r} \iiint \frac{a(x)}{g(x)} K\left(\frac{X_s - x}{h}\right) K\left(\frac{X_r - x}{h}\right) \\
&\quad \times Re\left[\phi_z(v,x)^* \varepsilon_{yz}(u,v,X_s)\varepsilon_y(u,X_r)^*\right] dW_1(u)dW_2(v)dx.
\end{aligned}$$

Proposition 7. *Under the conditions of Theorem 1,*

$$T_7 + T_8 + T_9 + T_{10} = o_P(1).$$

According to Proposition 1 to Proposition 7, we get the asymptotic mean B and the leading term U that determine the asymptotic distribution of our test:

$$\begin{aligned}
B &= B_1 + B_2 + B_3 + B_5 + B_6 \\
&= h^{-k/2} \iiint a(x)(1 - |\phi_y(u,x)|^2)(1 - |\phi_z(v,x)|^2) dW_1(u)dW_2(v)dx \int K^2(\tau)d\tau,
\end{aligned}$$

and

$$\begin{aligned}
U &= \tilde{U}_1 + \tilde{U}_2 + \tilde{U}_3 + \tilde{U}_4 + \tilde{U}_5 + \tilde{U}_6 \\
&= \frac{2}{nh^{3k/2}} \sum_{1 \le s < r \le n} U(\xi_s, \xi_r) \\
&= \frac{2}{nh^{3k/2}} \sum_{1 \le s < r \le n} \left[U_1(\xi_s,\xi_r) + U_2(\xi_s,\xi_r) + U_3(\xi_s,\xi_r) + U_4(\xi_s,\xi_r) + U_4(\xi_r,\xi_s)\right. \\
&\quad \left. + U_5(\xi_s,\xi_r) + U_5(\xi_r,\xi_s) + U_6(\xi_s,\xi_r) + U_6(\xi_r,\xi_s)\right].
\end{aligned}$$

The following Proposition presents the asymptotic property of the leading term U.

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

$$\begin{aligned}
V &= 2\int \left[\iint |\Phi_y(u_1+u_2,x)|^2 dW_1(u_1)dW_1(u_2) \iint |\Phi_z(v_1+v_2,x)|^2 dW_2(v_1)dW_2(v_2)\right] a^2(x)dx \\
&\quad \times \int \left[\int K(\tau)K(\tau+\eta)d\tau\right]^2 d\eta,
\end{aligned}$$

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 our test replaces the asymptotic mean B and variance V by their nonparametric estimators $\hat{B}$ and $\hat{V}$, which are given by Eqs. (17) and (18), we should show that $\hat{B}$ and $\hat{V}$ are consistent estimators for B and V, so that replacing B and V by $\hat{B}$ and $\hat{V}$ has asymptotic negligible impacts on the limiting distribution.

Proposition 9. *Under the conditions of Theorem 1, $\hat{B}$ and $\hat{V}$ are consistent estimators for B and V under $\mathbb{H}_0$ respectively.*

The proof of Theorem 1 will be completed provided Propositions 1 to 9 are proven, which we turn to next. Moreover, since the proofs of Propositions 1 to 7 are quite similar, for space, we only focus on the proofs of Proposition 1, Proposition 8, and Proposition 9.

Proof of Proposition 1. We first decompose T_1 as follows:

$$\begin{aligned} T_1 &= h^{k/2}\sum_{t=1}^{n}\int\left|\hat{\phi}_{yz}(u,v,X_t)-\phi_{yz}(u,v,X_t)\right|^2 a(X_t)dW_1(u)dW_2(v) \\ &= h^{k/2}\sum_{t=1}^{n}\int\left|\hat{\phi}_{yz}(u,v,X_t)-E\hat{\phi}_{yz}(u,v,X_t)\right|^2 a(X_t)dW_1(u)dW_2(v) \\ &\quad +2h^{k/2}\sum_{t=1}^{n}\int Re\left[(\hat{\phi}_{yz}(u,v,X_t)-E\hat{\phi}_{yz}(u,v,X_t))(E\hat{\phi}_{yz}(u,v,X_t)-\phi_{yz}(u,v,X_t))^*\right]a(X_t)dW_1(u)dW_2(v) \\ &\quad +h^{k/2}\sum_{t=1}^{n}\int\left|E\hat{\phi}_{yz}(u,v,X_t)-\phi_{yz}(u,v,X_t)\right|^2 a(X_t)dW_1(u)dW_2(v) \\ &= A_1+2R_1+R_2. \end{aligned} \tag{31}$$

Then the proof of Proposition 1 consists of the proofs of lemma 1 to lemma 3 below.

Lemma 1. *Under the conditions of Theorem 1,*

$$A_1 = B_1 + \tilde{U}_1 + o_P(1).$$

Lemma 2. *Let R_1 be defined as in (31), then $R_1 = o_P(1)$.*

Lemma 3. *Let R_2 be defined as in (31), then $R_2 = o_P(1)$.*

Proof of Lemma 1. The proof of Lemma 1 is quite similar as the proof of Proposition A.3 of Chen and Hong (2010). For space, we neglect it.

Proof of Lemma 2. Firstly, we decompose R_1 into two terms:

$$\begin{aligned} R_1 &= h^{k/2}\iint\sum_{t=1}^{n}Re\left[\sum_{s=1}^{n}\frac{a(X_t)}{nh^k g(X_t)}K\left(\frac{X_s-X_t}{h}\right)\left(\hat{\phi}_{yz}(u,v,X_s)-E\hat{\phi}_{yz}(u,v,X_s)\right)\right. \\ &\quad \left.\times\left(E\hat{\phi}_{yz}(u,v,X_t)-\phi_{yz}(u,v,X_t)\right)^*\right]dW_1(u)dW_2(v)\left[1+o_P(1)\right] \\ &= \frac{1}{nh^{k/2}}\sum_{t=1}^{n}\iint\frac{K(0)a(X_t)}{g(X_t)}Re\left[\left(\hat{\phi}_{yz}(u,v,X_t)-E\hat{\phi}_{yz}(u,v,X_t)\right)\left(E\hat{\phi}_{yz}(u,v,X_t)-\phi_{yz}(u,v,X_t)\right)^*\right] \\ &\quad \times dW_1(u)dW_2(v)\cdot[1+o_P(1)]+\frac{1}{nh^{k/2}}\sum_{t=1}^{n}\sum_{s\neq t}\iint\frac{a(X_t)}{g(X_t)}K\left(\frac{X_s-X_t}{h}\right) \\ &\quad \times Re\left[\left(\hat{\phi}_{yz}(u,v,X_s)-E\hat{\phi}_{yz}(u,v,X_s)\right)\left(E\hat{\phi}_{yz}(u,v,X_t)-\phi_{yz}(u,v,X_t)\right)^*\right]dW_1(u)dW_2(v)\cdot[1+o_P(1)] \\ &= [R_{11}+R_{12}]\cdot(1+o_P(1)). \end{aligned} \tag{32}$$

Then, we will prove $R_{11} = o_P(1)$.

Since

$$E\hat{\phi}_{yz}(u,v,x)-\phi_{yz}(u,v,x) = \frac{1}{2}h^2\nabla^2\phi(u,v,x)C_k + o_P(h^{2k}),$$

where $\nabla^2\phi(u,v,x) = \sum_{j=1}^{k}\frac{\partial^2}{\partial x_j^2}\phi(u,v,x)$ is the Laplacian of the function $\phi(u,v,x)$, then by Assumptions A.3 and A.4, we have $R_{11} = O_P(h^{3k/2}) = o_P(1)$ immediately.

Finally, let us prove $R_{12} = o_P(1)$.

Putting $\xi_t = (X_t, Y_t, Z_t)$, we define

$$\begin{aligned}\Psi(\xi_s,\xi_t) &= \iint \frac{a(X_t)}{g(X_t)} K\left(\frac{X_s - X_t}{h}\right) Re\left[\left(\hat{\phi}_{yz}(u,v,X_s) - E\hat{\phi}_{yz}(u,v,X_s)\right)\left(E\hat{\phi}_{yz}(u,v,X_t) - \phi_{yz}(u,v,X_t)\right)^*\right] dW_1(u)dW_2(v) \\ &+ \iint \frac{a(X_s)}{g(X_s)} K\left(\frac{X_t - X_s}{h}\right) Re\left[\left(\hat{\phi}_{yz}(u,v,X_t) - E\hat{\phi}_{yz}(u,v,X_t)\right)\left(E\hat{\phi}_{yz}(u,v,X_s) - \phi_{yz}(u,v,X_s)\right)^*\right] dW_1(u)dW_2(v),\end{aligned}$$

and

$$\begin{aligned}\Psi(\xi_s) &= \int \Psi(\xi_s,\xi_t)dP(\xi_t) \\ &= \iiint \frac{a(X_t)}{g(X_t)} K\left(\frac{X_s - X_t}{h}\right) Re\left[\left(\hat{\phi}_{yz}(u,v,X_s) - E\hat{\phi}_{yz}(u,v,X_s)\right)\left(E\hat{\phi}_{yz}(u,v,X_t) - \phi_{yz}(u,v,X_t)\right)^*\right] dW_1(u)dW_2(v)dX_t,\end{aligned}$$

where we have used the fact that $E\left[\left(\hat{\phi}_{yz}(u,v,X_t) - E\hat{\phi}_{yz}(u,v,X_t)\right)|X_t\right] = 0$. Then

$$\begin{aligned}R_{12} &= \frac{1}{nh^{k/2}} \sum_{1\le t<s\le n} [\Psi(\xi_s,\xi_t) - \Psi(\xi_s) - \Psi(\xi_t)] + \frac{2(n-1)}{nh^{k/2}} \sum_{t=1}^{n} \Psi(\xi_t) \\ &= R_{12}^{(1)} + R_{12}^{(2)}.\end{aligned} \tag{33}$$

Obviously, $E[\Psi(\xi_s,\xi_t) - \Psi(\xi_s) - \Psi(\xi_t)] = 0$, which means $E(R_{12}^{(1)}) = 0$. By Lemma A(ii) of Hjellvik et al. (1998), we have

$$\begin{aligned}\text{var}\left(R_{12}^{(1)}\right) &\le \frac{c}{n^2h^k} n^2 E\left[|\Psi(\xi_s,\xi_t) - \Psi(\xi_s) - \Psi(\xi_t)|^{2(1+\delta)}\right]^{\frac{1}{1+\delta}} \sum_{j=1}^{\infty} j^2 \beta^{\frac{\delta}{1+\delta}}(j) \\ &= O_P(h^{k+4}) \\ &= o_P(1).\end{aligned}$$

In addition,

$$\begin{aligned}\text{var}\left(R_{12}^{(2)}\right) &\le \frac{4(n-1)^2}{n^2h^k} \sum_{t=1}^{n} \text{var}(\Psi(\xi_t)) + \frac{4(n-1)^2}{n^2h^k} n \sum_{j=1}^{n-1}\left(1 - \frac{j}{n}\right) \text{cov}\left(\Psi(\xi_1), \Psi(\xi_{1+j})\right) \\ &\le cnh^{-k} O_P(h^{4+2k}) + 4nh^{-k} \sum_{j=1}^{\infty} j^2 \beta^{\frac{\delta}{1+\delta}}(j) O_P(h^{4+2k}) \\ &= O_P(nh^{4+k}) \\ &= o_P(1).\end{aligned}$$

Then, $R_{12}^{(1)} = o_P(1), R_{12}^{(2)} = o_P(1)$ follows from Chebyshev's inequality. Therefore, we finish the proof of Lemma 2.

Proof of Lemma 3. Define

$$\Upsilon(X_t) = \iint \left|E\hat{\phi}_{yz}(u,v,x_t) - \phi_{yz}(u,v,x_t)\right|^2 a(X_t)dW_1(u)dW_2(v).$$

Then

$$\begin{aligned}R_2 &= h^{k/2} \sum_{t=1}^{n} [\Upsilon(X_t) - E(\Upsilon(X_t))] + nh^{k/2} E[\Upsilon(X_t)] \\ &= R_2^{(1)} + R_2^{(2)}.\end{aligned}$$

Firstly, we prove $R_2^{(1)} = o_P(1)$.

$$\begin{aligned}\text{var}(R_2^{(1)}) &\leq h^k \sum_{t=1}^{n} \text{var}\left(\Upsilon(X_t)\right) + 2n \sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right) \text{cov}\left(\Upsilon(X_1)), \Upsilon(X_{1+j})\right)) \\ &\leq cnh^k O_P(h^8) + 2n\sum_{j=1}^{\infty} j^2 \beta^{\frac{\delta}{1+\delta}}(j) O_P(h^8) \\ &= O_P(nh^{8+k}) \\ &= o_P(1)\end{aligned}$$

Then, $R_2^{(1)} = o_P(1)$ follows from Chebyshev's inequality.

Besides, $R_2^{(2)}$ is a constant, satisfying:

$$\begin{aligned}R_2^{(2)} &= nh^{k/2} \iiint \left| \frac{1}{2}\nabla^2\phi(u,v,x)h^2 C_k + O_P(h^3)\right|^2 a(x)g(x)dW_1(u)dW_2(v)dx \\ &= O_P(nh^{k/2+4}) \\ &= B_1 \cdot o_P(1)\end{aligned}$$

which is a higher order term of B_1.

Proof of Proposition 8. To derive the asymptotic distribution, we apply Tenreiro's (1997) central limit theorem for degenerate U-statistics of a time series context process, which has been generally used by Su and White (2007, 2008), Hong and Lee (2013). Follow Tenreiro's (1997) central limit theorem, we have $\sigma_n^{-1}\sum_{1\leq s<r\leq n} U(\xi_s, \xi_r) \xrightarrow{d} N(0,1)$ if the following conditions are satisfied: 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 $\sigma_n^2 = \sum_{1\leq s<r\leq n} \text{var}[U(\xi_s,\xi_r)]$, and

$$\begin{aligned}u_n(p) &= \max\left\{\max_{1\leq i\leq n} \|U(\xi_i,\xi_1)\|_p, \|U(\xi_1,\bar{\xi}_1)\|_p\right\}, \\ v_n(p) &= \max\left\{\max_{1\leq i\leq n} \|G_{n1}(\xi_i,\xi_1)\|_p, \|G_{n1}(\xi_1,\bar{\xi}_1)\|_p\right\}, \\ w_n(p) &= \|G_{n1}(\xi_1,\xi_1)\|_p, \\ z_n(p) &= \max_{1\leq i\leq n}\max_{1\leq j\leq n}\left\{\|G_{nj}(\xi_i,\xi_1)\|_p, \|G_{nj}(\xi_1,\xi_i)\|_p, \|G_{nj}(\xi_1,\bar{\xi}_1)\|_p\right\}\end{aligned}$$

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

First, we calculate the asymptotic variance of $U(\xi_s,\xi_r)$

$$\sigma_0^2 = \text{var}\left[U(\xi_s,\xi_r)\right] = \iint U(\xi_s,\xi_r)^2 dP(\xi_s)dP(\xi_r).$$

Since $U(\xi_s,\xi_r)$ contains six terms, we need to calculate the variances of these six terms as well as their fifteen pairwise covariances. During the calculation, we have used the following facts: (1) The weighting functions $W_1(u), W_2(v)$ are symmetric, which means $\Phi_s(u_1+u_2,x) = \Phi_s(u_1-u_2,x)$; (2) Under the null hypothesis, Y_t is independent of Z_t conditional on X_t; (3) $\Phi_{yz}(u_1+u_2,v_1+v_2,x) \equiv \phi_{yz}(u_1+u_2,v_1+v_2,x) - \phi_{yz}(u_1,v_1,x)\phi_{yz}(u_2,v_2,x) = \phi_z(v_1+v_2,x)\Phi_y(u_1+u_2,x) + \phi_y(u_1+u_2,x)\Phi_z(v_1+v_2,x)$.

By tedious but straightforward algebra, we obtain

$$\begin{aligned}\sigma_0^2 &= h^{3k}\int a^2(x)\left[\iint |\Phi_y(u_1+u_2,x)|^2 dW_1(u_1)dW_1(u_2)\iint |\Phi_z(v_1+v_2,x)|^2 dW_2(v_1)dW_2(v_2)\right]dx \\ &\quad \times \int\left[\int K(\tau)K(\tau+\eta)d\tau\right]^2 d\eta\end{aligned}$$

Then, similar as Hjellvik et al. (1998), we obtain $\sigma_n^2 = \frac{n^2}{2}\sigma_0^2[1+o(1)]$. Thus, we have

$$V = \text{var}(U) = \frac{4}{n^2 h^{3k}}\sigma_n^2 = \frac{2}{h^{3k}}\sigma_0^2.$$

Now, we verify conditions (i)-(iv). Since $U(\xi_s, \xi_r)$ is the sum of six terms, $U(\xi_i, \eta)U(\xi_j, \tau)$ contains 36 terms, which is very tedious. But fortunately, these terms have the same order, and we can verify the first term $U_1(\xi_i, \eta)U_1(\xi_j, \tau)$ only.

$$\begin{aligned} E|U(\xi_i, \xi_1)|^p &\sim E\left|\iint \frac{a(x)}{g(x)} K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_1 - x}{h}\right) Re\left[\varepsilon_{yz}(u,v,x_i)\varepsilon_{yz}(u,v,x_1)^*\right] dW(u,v)dx\right|^p \\ &= h^{pk} E\left|\iint \frac{a(x_i - \tau h)}{g(x_i - \tau h)} K(\tau) K\left(\tau + \frac{x_1 - x_i}{h}\right) Re\left[\varepsilon_{yz}(u,v,x_i)\varepsilon_{yz}(u,v,x_1)^*\right] dW(u,v)d\tau\right|^p \\ &= O\left(h^{(p+1)k}\right) \end{aligned}$$

Therefore, we have $\|U(\xi_i, \xi_1)\|_p = O\left(h^{k+k/p}\right)$. By a similar argument, we can obtain the same order of magnitude for $\|U(\xi_1, \bar{\xi}_1)\|_p$. Hence, condition (i) holds for any $\delta_0 > 0$ and $\gamma_0 < \frac{1}{2}$.

Now, we verify condition (ii).

$$\begin{aligned} E\left|G_{n1}(\xi_i, \xi_1)\right|^p &= E\left|E_1\left[U(\xi_1, \xi_i)U(\xi_1, \xi_1)\right]\right|^p \\ &\sim E\left|\iiint \frac{a(x)}{g(x)} K\left(\frac{x_1 - x}{h}\right) K\left(\frac{x_i - x}{h}\right) Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_i)^*\right] dW(u,v)dx\right. \\ &\quad \left.\times \iint \frac{a(x')}{g(x')} K^2\left(\frac{x_1 - x'}{h}\right) Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_1)^*\right] dW(u,v)dx' dG(\xi_1)\right|^p \\ &= h^{2pk} E\left|\iiint \frac{a(x_1 - \tau h)}{g(x_1 - \tau h)} K(\tau) K\left(\tau + \frac{x_i - x_1}{h}\right) Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_i)^*\right] dW(u,v)d\tau\right. \\ &\quad \left.\times \iint \frac{a(x_1 - \eta h)}{g(x_1 - \eta h)} K^2(\eta)\, Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_1)^*\right] dW(u,v)d\eta dG(\xi_1)\right|^p \\ &= O\left(h^{3pk}\right) \end{aligned}$$

Thus, $\|G_{n1}(\xi_i, \xi_1)\|_p = O\left(h^{3k}\right)$. By a similar argument, we can obtain the same order of magnitude for $\|G_{n1}(\xi_1, \bar{\xi}_1)\|_p$. Consequently, condition (ii) is satisfied.

$$\begin{aligned} E\left|G_{n1}(\xi_1, \xi_1)\right|^p &= E\left|E_1\left[U(\xi_1, \xi_1)U(\xi_1, \xi_1)\right]\right|^p \\ &\sim E\left|\iiint \frac{a(x)}{g(x)} K^2\left(\frac{x_1 - x}{h}\right) Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_1)^*\right] dW(u,v)dx\right. \\ &\quad \left.\times \iint \frac{a(x')}{g(x')} K^2\left(\frac{x_1 - x'}{h}\right) Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_1)^*\right] dW(u,v)dx' dG(\xi_1)\right|^p \\ &= h^{2pk} E\left|\iiint \frac{a(x_1 - \tau h)}{g(x_1 - \tau h)} K^2(\tau)\, Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_1)^*\right] dW(u,v)d\tau\right. \\ &\quad \left.\times \iint \frac{a(x_1 - \eta h)}{g(x_1 - \eta h)} K^2(\eta)\, Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_1)^*\right] dW(u,v)d\eta dG(\xi_1)\right|^p \\ &= O\left(h^{2pk}\right) \end{aligned}$$

Thus, $w_n(p) = O\left(h^{2k}\right) = o(1)$, condition (iii) is satisfied.

$$\begin{aligned}
E|G_{nj}(\xi_i,\xi_1)|^p &= E\left|E_j\left[U(\xi_j,\xi_i)U(\xi_1,\xi_1)\right]\right|^p \\
&\sim E\left|\iiint \frac{a(x)}{g(x)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x}{h}\right)Re\left[\varepsilon_{yz}(u,v,x_j)\varepsilon_{yz}(u,v,x_i)^*\right]dW(u,v)dx\right. \\
&\quad \left.\times\iint \frac{a(x')}{g(x')}K^2\left(\frac{x_1-x'}{h}\right)Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_1)^*\right]dW(u,v)dx'dG(\xi_j)\right|^p \\
&= h^{2pk}E\left|\iiint \frac{a(x_j-\tau h)}{g(x_j-\tau h)}K(\tau)K\left(\tau+\frac{x_i-x_j}{h}\right)Re\left[\varepsilon_{yz}(u,v,x_j)\varepsilon_{yz}(u,v,x_i)^*\right]dW(u,v)d\tau\right. \\
&\quad \left.\times\iint \frac{a(x_1-\eta h)}{g(x_1-\eta h)}K^2(\eta)\,Re\left[\varepsilon_{yz}(u,v,x_1)\varepsilon_{yz}(u,v,x_1)^*\right]dW(u,v)d\eta dG(\xi_j)\right|^p \\
&= O\left(h^{3pk}\right)
\end{aligned}$$

By similar argument, we have $E|G_{nj}(\xi_1,\xi_i)|^p = O\left(h^{3pk+k}\right)$, $E|G_{nj}(\xi_1,\bar{\xi}_1)|^p = O\left(h^{3pk+k}\right)$. Therefore, $z_n(p) = O\left(h^{3k+k/p}\right)$ and condition (iv) is satisfied by setting $\gamma_1 = \frac{1}{\lambda}(3k+k/2) > 0$. Hence, we finish the proof of Proposition 8.

Proof of Proposition 9. We need to show $\hat{B}-B$ and $\hat{V}-V$ are higher order terms relative to B and V respectively. As $B = O_P(h^{-d_x/2})$ and $V = O_P(1)$, we should prove:

$$\hat{B}-B = h^{-d_x/2}o_P(1) \tag{34}$$

$$\hat{V}-V = O_P(1) \tag{35}$$

As the proofs of Eq. (34) and Eq. (35) are quite similar, we focus on the proof of Eq. (34).

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

According to Assumptions A.1-A.3, we just need to prove

$$|\hat{\phi}_y(u,x)|^2-|\phi_y(u,x)|^2 = o_P(1), \tag{36}$$

$$|\hat{\phi}_z(v,x)|^2-|\phi_z(v,x)|^2 = o_P(1). \tag{37}$$

To show (36), we first decompose $|\hat{\phi}_y(u,x)|^2-|\phi_y(u,x)|^2$ into two parts:

$$|\hat{\phi}_y(u,x)|^2-|\phi_y(u,x)|^2 = |\hat{\phi}_y(u,x)-\phi_y(u,x)|^2+2Re\left[\left(\hat{\phi}_y(u,x)-\phi_y(u,x)\right)\phi_y(u,x)^*\right].$$

According to Li and Racine (2006), we know

$$\begin{aligned}
|\hat{\phi}_y(u,x)-\phi_y(u,x)|^2 &= O_P\left(T^{-1}h^{-d_x}+h^4\right) = o_P(1), \\
\hat{\phi}_y(u,x)-\phi_y(u,x) &= O_P\left(T^{-1/2}h^{-d_x/2}+h^2\right) = o_P(1).
\end{aligned}$$

Since $\phi_y(u,x)$ is measurable, we obtain (36). The proof of (37) is quite similar as (36), so that we omit it. From (36) and (37), we can obtain $\hat{B}-B = h^{-d_x/2}o_P(1)$ immediately.

Proof of Theorem 2. Under the class of local alternatives $\mathbb{H}_1(a_n)$, we have

$$\begin{aligned}\hat{\sigma}_a(u,v,x) &= \hat{\phi}_{yz}(u,v,x) - \hat{\phi}_y(u,x)\hat{\phi}_z(v,x) \\ &= [\hat{\phi}_{yz}(u,v,x) - \phi_{yz}(u,v,x)] - \phi_z(v,x)[\hat{\phi}_y(u,x) - \phi_y(u,x)] - \phi_y(u,x)[\hat{\phi}_z(v,x) - \phi_z(v,x)] \\ &\quad -[\hat{\phi}_y(u,x) - \phi_y(u,x)][\hat{\phi}_z(v,x) - \phi_z(v,x)] + a_n\delta(u,v,x) \\ &= \hat{\sigma}(u,v,x) + a_n\delta(u,v,x)\end{aligned}$$

where $\hat{\sigma}(u,v,x)$ is given by Eq. (29).

Hence, our test statistic:

$$\begin{aligned}\hat{M}_a &= h^{k/2}\sum_{t=1}^{n}\iint |\hat{\sigma_a}(u,v,X_t)|^2 a(X_t)dW_1(u)dW_2(v) \\ &= \hat{M}_h + 2h^{k/2}\sum_{t=1}^{n}\iint Re\left[a_n\sigma(u,v,X_t)\delta(u,v,X_t)^*\right]a(X_t)dW_1(u)dW_2(v) \\ &\quad + h^{k/2}\sum_{t=1}^{n}\iint |a_n|^2|\delta(u,v,X_t)|^2 a(X_t)dW_1(u)dW_2(v) \\ &= \hat{M}_h + 2M_1 + M_2\end{aligned}$$

It is straightforward to show that $E(M_1) = O_P(n^{1/2}h^{(k+4)/2})\hat{B}^{1/2}$ and $\text{var}(M_1) = O_P(h^{k/2}(n^{-1}h^{-k} + h^4)) = o_P(1)$. Therefore $M_1 = o_P(1)$ by Chebyshev's inequality. Moreover, $M_2 \xrightarrow{p} \gamma$ by the Law of large number. In addition, under the class of local alternatives $\mathbb{H}_1(a_n)$, the asymptotic variance $V_a = \text{var}(\hat{M}_a) \xrightarrow{p} V$, as $M_1 = o_P(1)$ and $M_2 - \gamma = o_P(1)$. Consequently, we get the conclusion of Theorem 2.