UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 09-34
Economic Series (19)
May 2009

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

A Nonparametric Copula Based Test for Conditional Independence with Applications to Granger Causality[*]

Taoufik Bouezmarni[†] Jeroen V.K. Rombouts[‡] Abderrahim Taamouti[§]

May 30, 2009

Abstract

This paper proposes a new nonparametric test for conditional independence, which is based on the comparison of Bernstein copula densities using the Hellinger distance. The test is easy to implement because it does not involve a weighting function in the test statistic, and it can be applied in general settings since there is no restriction on the dimension of the data. In fact, to apply the test, only a bandwidth is needed for the nonparametric copula. We prove that the test statistic is asymptotically pivotal under the null hypothesis, establish local power properties, and motivate the validity of the bootstrap technique that we use in finite sample settings. A simulation study illustrates the good size and power properties of the test. We illustrate the empirical relevance of our test by focusing on Granger causality using financial time series data to test for nonlinear leverage versus volatility feedback effects and to test for causality between stock returns and trading volume. In a third application, we investigate Granger causality between macroeconomic variables.

JEL Classification: C12; C14; C15; C19; G1; G12; E3; E4; E52.

Keywords: Nonparametric tests; conditional independence; Granger non-causality; Bernstein density copula; bootstrap; finance; volatility asymmetry, leverage effect, volatility feedback effect; macroeconomics.

[*] We would like to thank Luc Bauwens, Christian Hafner, Todd Clark, Miguel Delgado, Jean-Marie Dufour, Jesús Gonzalo, Oliver Linton, Jean-Francois Richard, Roch Roy, Carlos Velasco, participants of the 2008 Canadian Econometric Study Group, 7th World Congress in Probability and Statistics, Joint Meeting of the Statistical Society of Canada and the Societe Francaise de Statistique, 2009 UC3M-LSE Workshop, and seminar participants in Lille3, KUL, Maastricht University, UCL, University of Pittsburgh and The Federal Reserve Bank of Kansas City for excellent comments that improved this paper. Financial support from the Spanish Ministry of Education through grants SEJ 2007-63098 is also acknowledged.
[†] Département de mathématiques, Université de Montréal. Address: Département de mathématiques et de statistique,Université de Montréal, C.P. 6128, succursale Centre-ville Montréal, Canada, H3C 3J7.
[‡] Institute of Applied Economics at HEC Montréal, CIRANO, CIRPEE, Université catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium. Address: 3000 Cote Sainte Catherine, Montréal (QC), Canada, H3T 2A7. TEL: +1-514 3406466; FAX: +1-514 3406469; e-mail:jeroen.rombouts@hec.ca.
[§] Economics Department, Universidad Carlos III de Madrid. Address: Departamento de Economía Universidad Carlos III de Madrid Calle Madrid, 126 28903 Getafe (Madrid) España. TEL: +34-91 6249863; FAX: +34-916249329; e-mail: ataamout@eco.uc3m.es.

1 Introduction

Testing in applied econometrics is often based on a parametric model that specifies the conditional distribution of the variables of interest. When the assumed parametric distribution is incorrectly specified, there is a risk of obtaining wrong conclusions with respect to a certain null hypothesis. Therefore, we would like to test the null hypothesis in a broader framework that allows us to leave free the specification of the underlying model. Nonparametric tests are well suited for this. In this paper, we propose a new nonparametric test for conditional independence between two random vectors of interest Y and Z, conditionally on a random vector X. The null hypothesis of conditional independence is defined when the density of Y conditional on Z and X is equal to the density of Y conditional only on X, almost everywhere.

We are particularly interested in Granger non-causality tests. Since Granger non-causality is a form of conditional independence, see Florens and Mouchart (1982), Florens and Fougère (1996) and Chalak and White (2008), these tests can be deduced from the conditional independence tests. The concept of causality introduced by Granger (1969) and Wiener (1956) is now a basic notion when studying the dynamic relationships between time series. This concept is defined in terms of predictability at horizon one of variable Y from its own past, the past of another variable Z, and possibly a vector X of auxiliary variables. Following Granger (1969), the causality from Z to Y one period ahead is defined as follows: Z causes Y if observations on Z up to time $t-1$ can help to predict Y at time t given the past of Y and X up to time $t-1$. Dufour and Renault (1998) generalize the concept of Granger causality by considering causality at a given horizon h and causality up to horizon h, where h is a positive integer and can be infinite. Such a generalization is motivated by the fact that, in the presence of auxiliary variables X, it is possible to have the variable Z not causing variable Y at horizon one, but causing it at a longer horizon $h > 1$. In this case, we have an indirect causality transmitted by the auxiliary variables X; see Sims (1980b), Hsiao (1982), and Lütkepohl (1993) for related work. More recently, White and Lu (2008) also extend Granger non-causality by introducing the notion of weak Granger non-causality and retrospective weak Granger non-causality. They analyze the relations between Granger non-causality and a concept of structural causality arising from a general non-separable recursive dynamic structural system.

To characterize and test Granger non-causality, it is common practice to specify linear parametric models. However, as noted by Baek and Brock (1992) the parametric linear Granger causality tests may have low power against certain nonlinear alternatives. Therefore, nonparametric regression tests and nonparametric independence and conditional independence tests have been proposed to deal with this issue. Nonparametric regression tests are introduced by Fan and Li (1996) who develop tests for the significance of a subset of regressors and tests for the specification of the semiparametric functional form of the regression function. Fan and Li (2001) compare the power properties of various kernel based nonparametric tests with the integrated conditional moment tests of Bierens and Ploberger (1997), and Delgado and Manteiga (2001) propose a test for selecting explanatory variables in nonparametric regression based on the bootstrap. Several nonparametric tests are also available to test for independence, including the rank based test of Hoeffding (1948), the empirical distribution based methods such as Blum, Kiefer, and Rosenblatt (1961) or Skaug and Tjostheim (1993), smoothing-based methods like Rosenblatt (1975), Robinson (1991), and Hong and White (2005).

The literature on nonparametric conditional independence tests is more recent. Linton and Gozalo (1997) develop a non-pivotal nonparametric empirical distribution function based test of conditional independence. The asymptotic null distribution of the test statistic is a functional of a Gaussian process and the critical values are computed using the bootstrap. Li, Maasoumi, and Racine (2009) propose a test designed for mixed discrete and continuous variables. They smooth both the discrete and continuous variables, with the smoothing parameters chosen via least-squares cross-validation. Their test has an asymptotic normal null distribution, however they suggest to use the bootstrap in finite sample settings. Lee and Whang (2009) provide a nonparametric test for the treatment effects conditional on covariates. They allow for both conditional average and conditional distributional treatment effects.

Few papers have been proposed to test for conditional independence using time series data. Su and White (2003) construct a class of smoothed empirical likelihood-based tests which are asymptotically normal under the null hypothesis, and derive their asymptotic distributions under a sequence of local alternatives. Their approach is based on testing distributional assumptions via an infinite collection of conditional moment restrictions, extending the finite unconditional and conditional moment tests of Kitamura (2001) and Tripathi and Kitamura (2003). The tests are shown to possess a weak optimality property in large samples and simulation results suggest that these tests behave well in finite samples. Su and White (2008) propose a nonparametric test based on kernel estimation of the density function and the weighted Hellinger distance. The test is consistent and asymptotically normal under β-mixing conditions. They use the nonparametric local smoothed bootstrap in finite sample settings. Su and White (2007), building on the previous test which uses densities, also propose a nonparametric test based on the conditional characteristic function. They work with the squared Euclidean distance, instead of the Hellinger distance, and need to specify two weighting functions in the test statistic.

In this paper, we propose a new approach to test for conditional independence. Our method is based on nonparametric copulas and the Hellinger distance. Copulas are a natural tool to test for conditional independence since they disentangle the dependence structure from the marginal distributions. They are usually parametric or semiparametric, see for example Chen and Fan (2006a) and Chen and Fan (2006b), though in the testing problem of this paper we prefer nonparametric copulas to give full weight to the data. To estimate nonparametrically the copulas, we use the Bernstein density copula. Using *i.i.d.* data, Sancetta and Satchell (2004) show that under some regularity conditions, any copula can be represented by a Bernstein copula. Bouezmarni, Rombouts, and Taamouti (2009) provide the asymptotic properties of the Bernstein density copula estimator using α-mixing dependent data. In this paper, under β-mixing conditions we show that our test statistic is asymptotically pivotal under the null hypothesis. To achieve this result, we subtract some bias terms from the Hellinger distance between the copula densities and then rescale by the proper variance. Furthermore, we establish local power properties and show the validity of the local smoothed bootstrap that we use in finite sample settings.

There are two important differences between our test and Su and White (2008)'s test. First, the total dimension d of the random vectors X, Y and Z in our nonparametric copula based test is not limited to be smaller than or equal to 7. Second, we do not need to select a weighting function to truncate the supports of continuous random variables which have support on the real line, because copulas are defined on the unit cube. In Su and White (2008), the choice of the weighting function is crucial for the properties of the test statistic. To apply our test, only a bandwidth is needed for the nonparametric copula. This is obviously appealing for the applied econometrician since the test becomes easy to implement. Other advantages are that the nonparametric Bernstein copula density estimates are guaranteed to be non-negative and therefore we avoid potential problems with the Hellinger distance. Furthermore, there is no boundary bias problem because, by smoothing with beta densities, the Bernstein density copula does not assign weight outside its support.

A simulation study reveals that our test has good finite sample size and power properties for a variety of typical data generating processes and different sample sizes. The empirical importance of testing for nonlinear causality is illustrated in three examples. In the first one, we examine the main explanations of the asymmetric volatility stylized fact using high-frequency data on S&P 500 Index futures contracts and find evidence of a nonlinear leverage effect and a nonlinear volatility feedback effect. In the second example, we study the relationship between stock index returns and trading volume. While both the linear and nonparametric tests find Granger causality from returns to volume, only the nonparametric test detects Granger causality from volume to returns. In the final example, we reexamine the causality between typical macroeconomic variables. The results show that linear Granger non-causality tests fail to detect the relationship between several of these variables, whereas our nonparametric tests confirm the statistical significance of these relationships.

The rest of the paper is organized as follows. The conditional independence test using the Hellinger distance and the Bernstein copula is introduced in Section 2. Section 3 provides the test statistic and its asymptotic properties. In Section 4, we investigate the finite sample size and power properties. Section 5 contains the three applications described above. Section 6 concludes. The proofs of the asymptotic results are presented in the Appendix.

2 Null hypothesis, Hellinger distance and the Bernstein copula

Let $\left\{(X_t', Y_t', Z_t')' \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3},\ t = 1, ..., T\right\}$ be a sample of stochastic processes in $\mathbb{R}^d$, where $d = d_1 + d_2 + d_3$, with joint distribution function F_{XYZ} and density function f_{XYZ}. We wish to test the conditional independence between Y and Z conditionally on X. Formally, the null hypothesis can be written in terms of densities as

$$H_0 : \mathsf{Pr}\left\{f_{Y|X,Z}(y \mid x, z) = f_{Y|X}(y \mid x)\right\} = 1,\ \forall y \in \mathbb{R}^{d_2}, \tag{1}$$

and the alternative hypothesis as

$$H_1 : \mathsf{Pr}\left\{f_{Y|X,Z}(y \mid x, z) = f_{Y|X}(y \mid x)\right\} < 1,\ \text{for some}\, y \in \mathbb{R}^{d_2},$$

where $f_{.|.}(\cdot|\cdot)$ denotes the conditional density. As we mentioned in the introduction, Granger non-causality is a form of conditional independence and to see that let us consider the following example. For $(Y, Z)'$ a Markov process of order 1, the null hypothesis which corresponds to Granger non-causality from Z to Y is given by

$$H_0 : \mathsf{Pr}\left\{f_{Y|X,Z}(y_t \mid y_{t-1},\ z_{t-1}) = f_{Y|X}(y_t \mid y_{t-1})\right\} = 1,$$

where in this case $y = y_t$, $x = y_{t-1}$, $z = z_{t-1}$ and $d_1 = d_2 = d_3 = 1$.

Next, we reformulate the null hypothesis (1) in terms of copulas. This will allow us to keep only the terms that involve the dependence among the random vectors. It is well known from Sklar (1959) that the distribution function of the joint process $(X', Y', Z')'$ can be expressed via a copula

$$F_{XYZ}(x, y, z) = C_{XYZ}\left(\bar{F}_X(x), \bar{F}_Y(y), \bar{F}_Z(z)\right), \tag{2}$$

where for simplicity of notation and to keep more space we denote $\bar{F}_X(x) = (F_{X_1}(x_1), ..., F_{X_{d_1}}(x_{d_1}))$, $\bar{F}_Y(y) = (F_{Y_1}(y_1), ..., F_{Y_{d_2}}(y_{d_2}))$, $\bar{F}_Z(z) = (F_{Z_1}(z_1), ..., F_{Z_{d_3}}(x_{d_3}))$, $F_{Q_i}(.)$, for $Q = X, Y, Z$, is the marginal distribution function of the *i-th* element of the vector Q, and $C_{XYZ}(.)$ is a copula function defined on $[0,1]^d$ which captures the dependence of $(X', Y', Z')'$. If we derive Equation (2) with respect to $(x', y', z')'$, we obtain the density function of the joint process $(X', Y', Z')'$ which can be expressed as

$$f_{XYZ}(x, y, z) = \prod_{j=1}^{d_1} f_{X_j}(x_j) \times \prod_{j=1}^{d_2} f_{Y_j}(y_j) \times \prod_{j=1}^{d_3} f_{Z_j}(z_j) \times c_{XYZ}\left(\bar{F}_X(x), \bar{F}_Y(y), \bar{F}_Z(z)\right), \tag{3}$$

where $f_{Q_j}(.)$, for $Q = X, Y, Z$, is the marginal density of the *j-th* element of the vector Q and $c_{XYZ}(.)$ is a copula density defined on $[0,1]^d$ of $(X', Y', Z')'$. Using Equation (3), we can show that the null hypothesis in (1) can be rewritten in terms of copula densities as

$$H_0 : \mathsf{Pr}\left\{c_{XYZ}\left(\bar{F}_X(x), \bar{F}_Y(y), \bar{F}_Z(z)\right) = c_{XY}\left(\bar{F}_X(x), \bar{F}_Y(y)\right) c_{XZ}\left(\bar{F}_X(x), \bar{F}_Z(z)\right)\right\} = 1,\ \forall y \in \mathbb{R}^{d_2} \tag{4}$$

against the alternative hypothesis

$$H_1 : \mathsf{Pr}\left\{c_{XYZ}\left(\bar{F}_X(x), \bar{F}_Y(y), \bar{F}_Z(z)\right) = c_{XY}\left(\bar{F}_X(x), \bar{F}_Y(y)\right) c_{XZ}\left(\bar{F}_X(x), \bar{F}_Z(z)\right)\right\} < 1, \quad \text{for some } y \in \mathbb{R}^{d_2},$$

where $c_{XY}(.)$ and $c_{XZ}(.)$ are the copula densities of the joint processes $(X', Y')'$ and $(X', Z')'$, respectively. Observe that under H_0, the dependence of the vector $(X', Y', Z')'$ is controlled by the dependence of $(X', Y')'$ and $(X', Z')'$ and not that of $(Y', Z')'$. Given the null hypothesis (4), our test statistic, say $H(c, C)$, is based on the Hellinger distance between $c_{XYZ}(u, v, w)$ and $c_{XY}(u, v)c_{XZ}(u, w)$, for $u \in [0, 1]^{d_1}$, $v \in [0, 1]^{d_2}$, $w \in [0, 1]^{d_3}$,

$$H(c, C) = \int_{[0,1]^d} \left(1 - \sqrt{\frac{c_{XY}(u, v)c_{XZ}(u, w)}{c_{XYZ}(u, v, w)}}\right)^2 dC_{XYZ}(u, v, w). \tag{5}$$

Under the null hypothesis, the measure $H(c, C)$ is equal to zero. The advantage of working with copulas instead of densities is that we integrate over $[0, 1]^d$ instead of $\mathbb{R}^d$. The Hellinger distance is often used for measuring the closeness between two densities and this is because it is simple to handle compared to L_∞ and L_1. Furthermore, it is symmetric and invariant to continuous monotonic transformations and it gives lower weight to outliers [see *e.g.* Beran (1977)]. The Hellinger distance in (5) can be estimated by

$$\begin{aligned}\hat{H} = H(\hat{c}, C_T) &= \int_{[0,1]^d} \left(1 - \sqrt{\frac{\hat{c}_{XY}(u, v)\hat{c}_{XZ}(u, v)}{\hat{c}_{XYZ}(u, v, w)}}\right)^2 dC_{XYZ,T}(u, v, w)\\ &= \frac{1}{T}\sum_{t=1}^{T} \left(1 - \sqrt{\frac{\hat{c}_{XY}\left(\bar{F}_X(X_t), \bar{F}_Y(Y_t)\right) \hat{c}_{XZ}\left(\bar{F}_X(X_t), \bar{F}_Z(Z_t)\right)}{\hat{c}_{XYZ}\left(\bar{F}_X(X_t), \bar{F}_Y(Y_t), \bar{F}_Z(Z_t)\right)}}\right)^2,\end{aligned}$$

where $\bar{F}_{X,T}(X_t)$, $\bar{F}_{Y,T}(Y_t)$ and $\bar{F}_{Z,T}(Z_t)$ with subscript T is to indicate the empirical analog of the distribution functions defined in $\bar{F}_X(X)$, $\bar{F}_Y(Y)$ and $\bar{F}_Z(Z)$, $C_{XYZ,T}(.)$ is the empirical copula defined by Deheuvels (1979), and $\hat{c}_{XY}(.)$, $\hat{c}_{XZ}(.)$ and $\hat{c}_{XYZ}(.)$ are the estimators of the copula densities $c_{XY}(.)$, $c_{XZ}(.)$ and $c_{XYZ}(.)$ respectively obtained using the Bernstein density copula defined below. Let us first set some additional notations. In what follows, we denote by

$$G_t = (G_{t1}, ..., G_{td}) = (\bar{F}_X(X_t), \bar{F}_Y(Y_t), \bar{F}_Z(Z_t)),$$

and its empirical analog

$$\hat{G}_t = (\hat{G}_{t1}, ..., \hat{G}_{td}) = (\bar{F}_{X,T}(X_t), \bar{F}_{Y,T}(Y_t), \bar{F}_{Z,T}(Z_t)).$$

The Bernstein density copula estimator of $c_{XYZ}(.)$ at a given value $g = (g_1, ..., g_d)$ is defined by

$$\hat{c}_{XYZ}(g_1, ..., g_d) = \hat{c}_{XYZ}(g) = \frac{1}{T}\sum_{t=1}^{T} K_k(g, \hat{G}_t), \tag{6}$$

where

$$K_k(g,\hat{G}_t) = k^d \sum_{\upsilon_1=0}^{k-1} ... \sum_{\upsilon_d=0}^{k-1} A_{\hat{G}_t,\upsilon} \prod_{j=1}^{d} p_{\upsilon_j}(g_j),$$

the integer k represent the bandwidth parameter, $p_{\upsilon_j}(g_j)$ is the binomial distribution

$$p_{\upsilon_j}(g_j) = \left(\begin{array}{c} k-1 \\ \upsilon_j \end{array} \right) g_j^{\upsilon_j} \, (1-g_j)^{k-\upsilon_j-1}, \text{ for } \upsilon_j = 0, \cdots, k-1,$$

and $A_{\hat{G}_t,\upsilon}$ is an indicator function

$$A_{\hat{G}_t,\upsilon} = \mathbf{1}_{\{\hat{G}_t \in B_\upsilon\}}, \text{ with } B_\upsilon = \left[\frac{\upsilon_1}{k}, \frac{\upsilon_1+1}{k}\right] \times ... \times \left[\frac{\upsilon_d}{k}, \frac{\upsilon_d+1}{k}\right].$$

The Bernstein estimators $\hat{c}_{XY}(.)$ and $\hat{c}_{XZ}(.)$ of $c_{XY}(.)$ and $c_{XZ}(.)$ respectively are defined in a similar way like for $\hat{c}_{XYZ}(.)$. Observe that the kernel $K_k(g,\hat{G}_t)$ can be rewritten as

$$K_k(g,\hat{G}_t) = \sum_{\upsilon_1=0}^{k-1} ... \sum_{\upsilon_d=0}^{k-1} A_{\hat{G}_t,\upsilon} \prod_{j=1}^{d} \mathcal{B}(x, \upsilon_j+1, k-\upsilon_j),$$

where $\mathcal{B}(x, \upsilon_j+1, k-\upsilon_j)$ is a beta density with shape parameters υ_j+1 and $k-\upsilon_j$ evaluated at x. $K_k(g,\hat{G}_t)$ can viewed as a smoother of the empirical density estimator by beta densities. The Bernstein density copula estimator in (6) is easy to implement, non-negative, integrates to one and is free from the boundary bias problem which often occurs with conventional nonparametric kernel estimators. Bouezmarni, Rombouts, and Taamouti (2009) establish the asymptotic bias, variance and the uniform almost convergence of Bernstein density copula estimator for α-mixing data. These properties are necessary to prove the asymptotic normality of our test statistic. Notice that some other nonparametric copula density estimators are proposed in the literature. For example, Gijbels and Mielniczuk (1990) suggest nonparametric kernel methods and use the reflection method to overcome the boundary bias problem, and more recently Chen and Huang (2007) use the local linear estimator. Fermanian and Scaillet (2003) derive the asymptotic properties of kernel estimators of nonparametric copulas and their derivatives in the context of time series data.

In the next section, we derive the asymptotic normality of our test statistic $\hat{H}$ under the null hypothesis of conditional independence. A few bias terms and a standardization are required to obtain a pivotal test statistic that converges to the standard normal distribution. We also establish the local power properties of the test and show the validity of the local smoothed bootstrap procedure.

3 Asymptotic distribution and power of the test statistic

Since we are interested in time series data, we need to specify the dependence in the process of interest. In what follows, we consider β-mixing dependent variables. The β-mixing condition is required to show the asymptotic normality of U-statistics as our test statistic; see Tenreiro (1997) and Fan and Li (1999) among others. To establish the asymptotic normality of the test statistic $\hat{H}$, we also need to apply the results of Bouezmarni, Rombouts, and Taamouti (2009). The latter are valid under weak condition of α-mixing processes. However, no asymptotic normality for U-statistics seems to be available under α-mixing dependence. Now let us recall the definition of a β-mixing process. For $\{\mathcal{W}_t = (X_t', Y_t', Z_t')'\,;\, t \geq 0\}$ a strictly stationary stochastic process and $\mathcal{F}_t^s$ a sigma algebra generated by $(\mathcal{W}_s, ..., \mathcal{W}_t\,)$ for $s \leq t$, the process $\mathcal{W}$ is called β-mixing or absolutely regular, if

$$\beta(l) = \sup_{s\in\mathbb{N}} \mathbb{E}\left[\sup_{\mathcal{A}\in\mathcal{F}_{s+l}^{+\infty}} \left|P(\mathcal{A}|\mathcal{F}_{-\infty}^{s}) - P(\mathcal{A})\right|\right] \to 0, \quad \text{as} \quad l \to \infty.$$

To prove the asymptotic normality of our test statistic, additional regularity assumptions are needed. We consider a set of standard assumptions on the stochastic process and bandwidth parameter of the Bernstein copula density estimator.

Assumptions on the stochastic process

(A1.1) $\left\{(X_t', Y_t', Z_t')' \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3} \equiv \mathbb{R}^d,\ t \geq 0\right\}$ is a strictly stationary β-mixing process with coefficient $\beta_l = O(\rho^l)$, for some $0 < \rho < 1$.

(A1.2) G_t has a copula function C_{XYZ} and copula density c_{XYZ}. We assume that c_{XYZ} is twice continuously differentiable and bounded away from zero, *i.e.*, $\inf_{g\in[0,1]^d}\{c_{XYZ}(g)\} > 0$.

Assumptions on the bandwidth parameter

(A1.3) We assume that for $k \to \infty$, $T\,k^{-(d/2)-2} \to 0$ and $T^{-1/2}k^{d/4}\ln(T) \to 0$.

Assumption (**A1.1**), is satisfied by many processes, such as ARMA and ARCH processes, as documented for example by Carrasco and Chen (2002) and Meitz and Saikkonen (2002). This assumption is required to establish the central limit theorem of U-statistics for dependent data. In Assumption (**A1.2**), the second differentiability of c_{XYZ} is required by Bouezmarni, Rombouts, and Taamouti (2009) in order to calculate the bias of the Bernstein copula estimator. Further, since we use Hellinger distance, the copula should be positive, *i.e.* $\inf_{g\in[0,1]^d}\{c_{XYZ}(g)\} > 0$. Assumption (**A1.3**) is needed to cancel out a bias term in the test statistic and for the almost sure convergence of the Bernstein copula estimator. The bandwidth parameter k plays the inverse role compared to that of the standard nonparametric kernel, that is a large value of k reduces the bias but increases the variance. If we choose $k = O(T^{\xi})$, then ξ should be in $(2/(d+4), 2/d)$ in order to satisfy Assumption (**A1.3**). We now state the asymptotic distribution of our test statistic under the null hypothesis.

Theorem 1 *Under assumptions **(A1.1)-(A1.3)** and H_0, we have*

$$BRT = \frac{T\, k^{-d/2}}{\sigma}\left(4\hat{H} - C_1 T^{-1} k^{d/2} - \hat{B}_1 T^{-1} k^{(d_1+d_2)/2} - \hat{B}_2 T^{-1} k^{(d_1+d_3)/2} - \hat{B}_3 T^{-1} k^{d_1/2}\right) \to \mathcal{N}(0,1),$$

where $C_1 = 2^{-d}\pi^{d/2}$, $\sigma = \sqrt{2}\,(\pi/4)^{d/2}$ and

$$\hat{B}_1 = -2^{-(d_1+d_2-1)}\pi^{(d_1+d_2)/2} + T^{-1}\sum_{t=1}^{T} \frac{\prod_{j=1}^{d_1+d_2}\left(4\pi\,\hat{G}_{tj}(1-\hat{G}_{tj})\right)^{-1/2}}{\hat{c}_{XY}(\hat{G}_{t1},\ldots,\hat{G}_{t(d_1+d_2)})},$$

$$\hat{B}_2 = -2^{-(d_1+d_3-1)}\pi^{(d_1+d_3)/2} + T^{-1}\sum_{t=1}^{T} \frac{\prod_{j=1}^{d_1+d_3}\left(4\pi\,\hat{G}_{tj}(1-\hat{G}_{tj})\right)^{-1/2}}{\hat{c}_{XZ}(\hat{G}_{t1},\ldots,\hat{G}_{td_1},\hat{G}_{t(d_1+d_2+1)},\ldots,\hat{G}_{td})},$$

$$\hat{B}_3 = 2^{-(d_1-1)}\pi^{-(d_1/2)} T^{-1}\sum_{t=1}^{T} \frac{\hat{c}_X(\hat{G}_{t1},\ldots,\hat{G}_{td_1})}{\sqrt{\prod_{j=1}^{d_1}\hat{G}_{tj}(1-\hat{G}_{tj})}}.$$

$\hat{c}_X(.)$ is the Bernstein density copula estimator of the copula density $c_X(.)$ of the vector X. For a given significance level α, we reject the null hypothesis when $BRT > z_\alpha$, where z_α is the critical value from the standard normal distribution.

■

Note that the above asymptotic normality of the test statistic BRT does not require a limitation on the dimension d of the vector $(X', Y', Z')'$. Interestingly, in the typical case when $d_1 = 1$, the bias correction term $\hat{B}_3$ does not have to be estimated since $\hat{c}_X(\hat{G}_{t1}) = 1$ and the remaining sum over the T observations in the denominator is constant. Furthermore, the variance σ^2 does not have to be estimated from the data, it only depends on d. For the bias correction terms and in comparison with Su and White (2008), our test statistic does not require additional estimators that bring in extra assumptions.

Now, to evaluate the power of the proposed test, we consider the following sequence of local alternatives

$$H_1(\alpha_T): \quad f^{[T]}(y|x,z) = f^{[T]}(y|x)\left\{1 + \alpha_T\Delta(x,y,z) + o(\alpha_T)\Delta_T(x,y,z)\right\}, \tag{7}$$

where $f^{[T]}(y|x,z)$ (*resp.* $f^{[T]}(y|x)$) is the conditional density of $Y_{T,t}$ given $X_{T,t}$ and $Z_{T,t}$ (*resp.* of $Y_{T,t}$ given $X_{T,t}$). The process $\left\{(X'_{T,t}, Y'_{T,t}, Z'_{T,t})', \in \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\times\mathbb{R}^{d_3} \equiv \mathbb{R}^d, \text{ for } t = 1,..,T \text{ and } T \geq 1\right\}$ is assumed to be a strictly stationary β-mixing process with coefficient β_k^T such that $\sup_T \beta_k^T = O(\rho^k)$, for some $0 < \rho < 1$ and $\alpha_T \to 0$ as $T \to \infty$. The functions Δ and Δ_T satisfy the power assumptions below. The local alternatives in (7) are also considered by Gouriéroux and Tenreiro (2001). Similarly, power properties for other alternatives like Horowitz and Spokoiny (2001) can also be computed without any problem. The following additional assumptions are needed to establish the power properties of our test.

Power assumptions

(A2.1) $1 + \alpha_T \Delta(x, y, z) + o(\alpha_T)\Delta_T(x, y, z) \geq 0$, for all $(x', y', z')' \in \mathbb{R}^d$ and all $T \in \mathbb{N}$.

(A2.2) $\int \Delta(x, y, z) f^{[T]}(x, y) f^{[T]}(z|x) d(x, y, z) = \int \Delta_T(x, y, z) f^{[T]}(x, y) f^{[T]}(z|x) d(x, y, z) = 0$, for all $T \in \mathbb{N}$.

(A2.3) $\int |\Delta(x, y, z)|^2 f^{[T]}(x, y, z) d(x, y, z)$ and $\int |\Delta_T(x, y, z)|^2 f^{[T]}(x, y, z) d(x, y, z)$ are finite, for all $T \in \mathbb{N}$.

(A2.4) $\lim_T c_{XYZ}^{[T]}(u, v, w) = c_{XYZ}(u, v, w)$, where $c^{[T]}$ is the copula density of $(X'_{T,t}, Y'_{T,t}, Z'_{T,t})'$.

Assumption (**A2.1**) guarantees the positivity of $f^{[T]}(x, y, z)$ and assumption (**A2.2**) ensures that its integral is equal to one. Assumption (**A2.2**) is important for the proof of Lemma 1 in the Appendix. Next, we state the power function of our test.

Proposition 1 *Under assumptions **(A1.1)**-**(A1.3)** and **(A2.1)**-**(A2.4)**, and for $\alpha_T = T^{-1/2} k^{-d/4}$, if $H_1(\alpha_T)$ holds then we have*

$$BRT \to \mathcal{N}\left(\frac{1}{\sigma} \int \Delta^2 \left(\bar{F}_X^{-1}(u), \bar{F}_Y^{-1}(v), \bar{F}_Z^{-1}(w)\right) dC_{XYZ}(u, v, w),\ 1\right),$$

where

$$\bar{F}_X^{-1}(u) = \left(F_{X_1}^{-1}(u_1), ..., F_{X_{d_1}}^{-1}(u_{d_1})\right),$$

$$\bar{F}_Y^{-1}(v) = \left(F_{Y_1}^{-1}(v_1), ..., F_{Y_{d_2}}^{-1}(v_{d_2})\right),$$

$$\bar{F}_Z^{-1}(w) = \left(F_{Z_1}^{-1}(w_1), ..., F_{Z_{d_3}}^{-1}(w_{d_3})\right),$$

and $F_{Q_i}^{-1}(.)$, for $Q = X, Y, Z$, is the inverse distribution function of the i-th element of the vector Q. Hence, the power of the test based on the Bernstein density copula estimator is asymptotically $1 - \Phi\left(z_\alpha - \frac{1}{\sigma} \int \Delta^2 \left(\bar{F}_X^{-1}(u), \bar{F}_Y^{-1}(v), \bar{F}_Z^{-1}(w)\right) dC_{XYZ}(u, v, w)\right)$, where $\Phi(.)$ is the standard normal distribution function and z_α is the critical value at significance level α.

■

The above results on the distribution of the test statistic are valid only asymptotically. For finite samples, the bootstrap is used to compute the *p-values.* A simple bootstrap, *i.e.* resampling from the empirical distribution, will not conserve the conditional dependence structure in the data and hence sampling under the null hypothesis is not guaranteed. To prevent this from occurring, we use the local smoothed bootstrap suggested by Paparoditis and Politis (2000). The method is easy to implement in the following five steps: **(1)** we draw the sample X_t^* from the nonparametric kernel estimator $T^{-1} h^{-d_1} \sum_{t=1}^T L(X_t - x)/h$; **(2)** conditional on X_t^*, we draw Y_t^* and Z_t^* independently from the conditional density, that is $h^{-d_2} \sum_{t=1}^T L\left((X_t - x)/h\right) L\left((Y_t - y)/h\right) / \sum_{t=1}^T L\left((X_t - x)/h\right)$ and $h^{-d_2} \sum_{t=1}^T L\left((X_t - x)/h\right) L\left((Z_t - y)/h\right) / \sum_{t=1}^T K\left((X_t - x)/h\right)$, respectively; **(3)** based on the bootstrap sample, we compute the bootstrap statistic BRT^* in the same way as BRT; **(4)** we repeat the steps (1)-(3) B times so that we obtain BRT_j^*, for $j = 1, ..., B$; **(5)** the bootstrap *p-value* is computed as $p^* = B^{-1} \sum_{j=1}^{B} 1_{\{BRT_j^* > BRT\}}$. For given significance level α, we reject the null hypothesis if $p^* < \alpha$. To achieve the validity of the local bootstrap for the conditional independence test using the Bernstein copula estimator, we consider additional assumptions on the kernel K and bandwidth h.

Assumptions on bootstrap kernel and bandwidth

(A3.1) The kernel L is a product kernel of a bounded symmetric kernel density $l : \mathbb{R} \to \mathbb{R}^+$ such that $\int l(u)du = 1$ and $\int ul(u)du = 0$.

(A3.2) l is r times continuously differentiable such that $\int u^j l^{(r)}(u)du = 0$ for $j = 0, ..., r-1$ and $\int u^r l^{(r)}(u)du < \infty$, where $l^{(r)}$ is the rth-derivative of l.

(A3.3) As $T \to \infty$, $h \to 0$, and $Th^{d+2r}/(\ln T)^\gamma \to C > 0$, for some $\gamma > 0$.

Under Assumptions **(A3.1)**-**(A3.3)**, the almost sure convergence of the smoothed kernel estimators is fulfilled; see Paparoditis and Politis (2000). The following proposition states the validity of the bootstrap.

Proposition 2 *Under assumptions **(A1.1)**-**(A1.3)** and **(A3.1)**-**(A3.3)**, we have*

$$BRT^* \to \mathcal{N}(0, 1).$$

■

4 Finite sample size and power properties

In this section, we study the performance of the BRT test in a finite sample setting. To implement the Bernstein density copula estimator in the simulations and applications, we define $k^{*t} = (k_1^{*t}, ..., k_d^{*t}) = [k\hat{G}_t]$, where $[.]$ denotes the integer part of each element, from which we have

$$K_k(g, \hat{G}_t) = k^d \prod_{j=1}^{d} p_{k_j^{*t}}(g_j).$$

The data generating processes (DGP's) are detailed in Table 1. The first four DGP's simulate data that allow to illustrate the size properties of the tests: DGP3s includes the ARCH model of Engle (1982) and the DGP4s GARCH model of Bollerslev (1986). In the last six DGP's, the null hypothesis of conditional independence is not true and therefore serve to illustrate the power of the tests: DGP1p to DGP3p exhibit linear and nonlinear causality in the conditional mean and DGP4p to DGP6p nonlinear causality through the conditional variance.

Table 1: Data generating processes used in the simulations

DGP	X_t	Y_t	Z_t
1s	ε_{1t}	ε_{2t}	ε_{3t}
2s	Y_{t-1}	$Y_t = 0.5Y_{t-1} + \varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
3s	Y_{t-1}	$Y_t = (0.01 + 0.5Y_{t-1}^2)^{0.5}\varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
4s	Y_{t-1}	$Y_t = \sqrt{h_{1,t}}\varepsilon_{1t}$ $h_{1,t} = 0.01 + 0.5Y_{t-1}^2 + 0.9h_{1,t-1}$	$Z_t = \sqrt{h_{2,t}}\varepsilon_{2t}$ $h_{2,t} = 0.01 + 0.5Z_{t-1}^2 + 0.9h_{2,t-1}$
1p	Y_{t-1}	$Y_t = 0.5Y_{t-1} + 0.5Z_{t-1} + \varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
2p	Y_{t-1}	$Y_t = 0.5Y_{t-1} + 0.5Z_{t-1}^2 + \varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
3p	Y_{t-1}	$Y_t = 0.5Y_{t-1}Z_{t-1}^2 + \varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
4p	Y_{t-1}	$Y_t = 0.5Y_{t-1} + Z_{t-1}\varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
5p	Y_{t-1}	$Y_t = \sqrt{h_{1,t}}\varepsilon_{1t}$ $h_{1,t} = 0.01 + 0.5Y_{t-1}^2 + 0.25Z_{t-1}^2$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
6p	Y_{t-1}	$Y_t = \sqrt{h_{1,t}}\varepsilon_{1t}$ $h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4Y_{t-1}^2 + 0.5Z_{t-1}^2$	$Z_t = \sqrt{h_{2,t}}\varepsilon_{2t}$ $h_{2,t} = 0.01 + 0.5Z_{t-1}^2 + 0.9h_{2,t-1}$

We simulate (X_t, Y_t, Z_{t-1}), $t = 1, \ldots, T$. $(\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})' \sim N(0, I_3)$ and i.i.d.

As we explained before, our test statistic is asymptotically normally distributed under the null hypothesis, though we will use the local smooth bootstrap to approximate its finite sample distribution. The BRT test depends on the bandwidth k to estimate the copula densities. We take k the integer part of $cT^{1/2}$ for $c = 1, 1.5, 2$. We consider various values of k to evaluate the sensitivity with respect to the test results. This is common practice in nonparametric testing where no optimal bandwidth is available. To keep the computing time in our simulation study reasonable, we consider the sample sizes $T = 200$ and $T = 300$, $B = 200$ bootstrap replications with resampling bandwidths chosen by the standard rule of thumb. We use 250 iterations to compute the empirical size and power. As a comparison, we also include the linear Granger non-causality test in the Monte Carlo experiment to appreciate the loss of power against nonlinear alternatives. This simply tests if Z_{t-1} should enter the regression of Y_t on Y_{t-1}.

The size properties of the tests are given in Table 2. The linear test, LIN, behaves well as expected since the null hypothesis of conditional independence is true. The BRT test tends to be slightly conservative in some situations. For DGP2s with sample size 200 and $c = 1$, the sizes are 2.8% instead of 5% and 8.4% instead of 10%. As expected, we also see that the realized size varies with the bandwidth k, in the majority of the cases the relation is positive.

Table 2: Size properties of the tests

	DGP1s	DGP2s	DGP3s	DGP4s	DGP1s	DGP2s	DGP3s	DGP4s
		$T = 200$, $\alpha = 5\%$				$T = 200$, $\alpha = 10\%$		
LIN	0.043	0.053	0.042	0.050	0.084	0.109	0.101	0.090
BRT, c=1	0.040	0.028	0.028	0.040	0.080	0.084	0.084	0.100
BRT, c=1.5	0.048	0.032	0.028	0.044	0.100	0.056	0.048	0.108
BRT, c=2	0.052	0.028	0.032	0.072	0.112	0.056	0.068	0.140
		$T = 300$, $\alpha = 5\%$				$T = 300$, $\alpha = 10\%$		
LIN	0.049	0.057	0.054	0.050	0.102	0.109	0.114	0.108
BRT, c=1	0.040	0.032	0.036	0.052	0.076	0.076	0.064	0.084
BRT, c=1.5	0.036	0.028	0.024	0.052	0.080	0.072	0.056	0.104
BRT, c=2	0.048	0.024	0.020	0.092	0.128	0.052	0.044	0.124

Empirical size for a test at the α level based on 250 replications. The number of bootstrap resamples is B=200. LIN means linear test and BRT our test. The bandwidth k is the integer part of $cT^{1/2}$.

The power properties of the tests are presented in Table 3. We observe that the linear test has only excellent power to detect linear Granger causality. In fact, the power is 1 in DGP1p. For the other DGP's which involve nonlinear dependence, the linear test fails to achieve considerable power. The BRT tests have high power for all DGP's and the rise in its power from sample size 200 to 300 is important. Note also that generally the power of the BRT tests goes down with c, which is not surprising since the bandwidth k in the BRT test plays the inverse role of a kernel bandwidth in nonparametric density estimation.

Table 3: Power properties of the tests

	DGP1p	DGP2p	DGP3p	DGP4p	DGP5p	DGP6p
$T = 200$, $\alpha = 5\%$						
LIN	0.999	0.337	0.213	0.126	0.163	0.153
BRT, c=1	0.996	0.996	0.972	0.984	0.888	0.772
BRT, c=1.5	0.984	0.992	0.940	0.996	0.936	0.740
BRT, c=2	0.936	0.976	0.864	0.996	0.912	0.684
$T = 300$, $\alpha = 5\%$						
LIN	1.000	0.354	0.250	0.113	0.172	0.143
BRT, c=1	1.000	1.000	0.992	0.996	0.980	0.928
BRT, c=1.5	1.000	1.000	0.992	1.000	0.988	0.928
BRT, c=2	0.988	1.000	0.972	1.000	0.992	0.876
$T = 200$, $\alpha = 10\%$						
LIN	1.000	0.436	0.284	0.175	0.239	0.233
BRT, c=1	0.996	1.000	0.828	0.988	0.932	0.860
BRT, c=1.5	0.992	0.996	0.956	0.996	0.976	0.824
BRT, c=2	0.968	0.988	0.908	0.996	0.952	0.800
$T = 300$, $\alpha = 10\%$						
LIN	1.000	0.442	0.327	0.176	0.253	0.209
BRT, c=1	1.000	1.000	0.992	0.996	0.984	0.932
BRT, c=1.5	1.000	1.000	0.996	1.000	0.992	0.960
BRT, c=2	0.992	1.000	0.988	1.000	0.996	0.928

Empirical power for a test at the α level based on 250 replications. The number of bootstrap resamples is B=200. LIN means linear test and BRT our test. The bandwidth k is the integer part of $cT^{1/2}$.

5 Empirical applications

In this section, we consider three empirical applications to illustrate the importance of testing for nonlinear causality and the usefulness of our nonparametric test in this context. In the first example, we use high-frequency equity index data to analyze the main explanations of the asymmetric volatility stylized fact. In the second example, we study the causality between stock index returns and trading volume. In the final example, we reexamine the causality between monetary policy and the real economy.

5.1 Application 1: Nonlinear volatility feedback effect

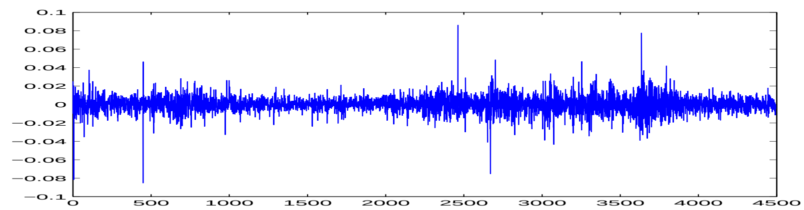
One of the many stylized facts about equity returns is an asymmetric relationship between returns and volatility (hereafter asymmetric volatility): volatility tends to rise following negative returns and fall following positive returns. The literature has two explanations for the asymmetric volatility. The first one is the leverage effect and means that a decrease in the price of an asset increases financial leverage and the probability of bankruptcy, making the asset riskier, hence an increase in volatility, see Black (1976) and Christie (1982). The second explanation is the volatility feedback effect which is related to the time-varying risk premium theory: if volatility is priced, an anticipated increase in volatility would raise the rate of return, requiring an immediate stock price decline in order to allow for higher future returns, see Pindyck (1984), French and Stambaugh (1987), and Campbell and Hentschel (1992), among others.

Empirically, studies focusing on the leverage hypothesis, see Christie (1982) and Schwert (1989), conclude that it cannot completely account for changes in volatility. For the volatility feedback effect, there are conflicting empirical findings. French and Stambaugh (1987) and Campbell and Hentschel (1992) find a positive relation between volatility and expected returns, while Turner, Startz, and Nelson (1989), Glosten and Runkle (1993), and Nelson (1991) find the relation to be negative but statistically insignificant. Using high-frequency data, Dufour, Garcia, and Taamouti (2008) measure a strong dynamic leverage effect for the first three days, whereas the volatility feedback effect is found to be insignificant at all horizons [see also Bollerslev, Litvinova, and Tauchen (2006)].

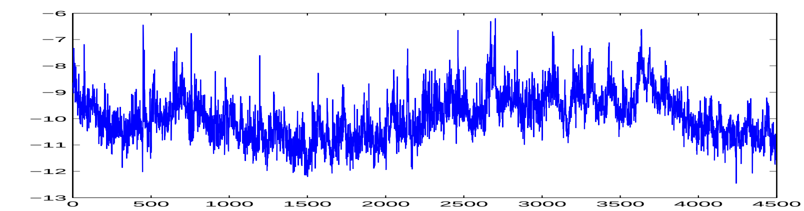
5.1.1 Data description

We consider tick-by-tick transaction prices for the S&P 500 Index futures contracts traded on the Chicago Mercantile Exchange, over the period January 1988 to December 2005 (4494 trading days). Following Huang and Tauchen (2005), we eliminate a few days where trading was thin and the market was only open for a shortened session. Due to the unusually high volatility at the opening of the market, we omit the first five minutes of each trading day, see Bollerslev, Litvinova, and Tauchen (2006). We compute the continuously compounded returns over each five-minute interval by taking the difference between the logarithm of the two tick prices immediately preceding each five-minute mark, implying 77 observations per day. Because volatility is latent, it is approximated by either realized volatility or bipower variation. Daily realized volatility is defined as the summation of the corresponding high-frequency intradaily squared returns $RV_{t+1} =$

$\sum_{j=1}^{h} r^2_{(t+j\Delta,\Delta)}$, where $r^2_{(t+j\Delta,\Delta)}$ are the discretely sampled Δ-period returns. Properties of realized volatility are provided by Andersen, Bollerslev, and Diebold (2003) [see also Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2002a), Barndorff-Nielsen and Shephard (2002b) and Comte and Renault (1998)]. The bipower variation is given by sum of cross product of the absolute value of intradaily returns $BV_{t+1} = \frac{\pi}{2}\sum_{j=2}^{h} \mid r_{(t+j\Delta,\Delta)} \mid\mid r_{(t+(j-1)\Delta,\Delta)} \mid$ .Its properties are provided by Barndorff-Nielsen and Shephard (2003) [see also Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2005)]. The sample paths for the returns and realized volatility are displayed in Figure 1.



(a) Returns



(b) Realized volatility (in logarithms)

Figure 1: S&P 500 futures daily data series

5.1.2 Causality tests

To test for linear causality, we estimate a first order vector autoregressive model (VAR(1)). This yields the following results [t-statistics are between brackets]

$$\begin{bmatrix} \hat{r}_t \\ \widehat{\ln(RV_t)} \end{bmatrix} = \begin{bmatrix} \underset{[0.982]}{0.001473} \\ \underset{[-24.670]}{-2.342446} \end{bmatrix} + \begin{bmatrix} \underset{[-2.8974]}{-0.043375} & \underset{[0.995]}{0.000150} \\ \underset{[-6.3418]}{-6.000874} & \underset{[80.178]}{0.764097} \end{bmatrix} \begin{bmatrix} r_{t-1} \\ \ln(RV_{t-1}) \end{bmatrix} \quad \begin{matrix} R^2 = 0.002 \\ R^2 = 0.597 \end{matrix} . \tag{8}$$

The results of linear causality tests between returns and volatility are presented in Table 4 [see also Equation 8]. We find convincing evidence that return causes volatility. However, given the *p-value* of 0.320 we find that there is no impact (linear causality) from volatility to return. Consequently, we conclude that there is a leverage effect but not a volatility feedback effect. Considering different orders for vector autoregressive model leads to the same conclusion. Further, replacing realized volatility ($\ln(RV_t)$) with bipower variation ($\ln(BV_t)$) also yields similar results.

To test for the presence of nonlinear volatility feedback and leverage effects, we consider the following null hypotheses: $H_0 : f(r_t \mid r_{t-1},\ \ln(RV_{t-1})) = f(r_t|r_{t-1})$ and $H_0 : f(\ln(RV_t) \mid \ln(RV_{t-1}),\ r_{t-1}) = f(\ln(RV_t)|\ln(RV_{t-1}))$, respectively. The results are presented in Table 4. At a

Table 4: P-values for linear and nonlinear causality tests

Test statistic / H_0	No feedback	No leverage
LIN	0.320	0.000
BRT, $c = 1$	0.000	0.000
BRT, $c = 1.5$	0.000	0.000
BRT, $c = 2$	0.020	0.000

Linear and Nonlinear causality tests between returns (r) and volatility (approximated by $\ln(RV)$). LIN and BRT correspond to linear test and our nonparametric test, respectively.

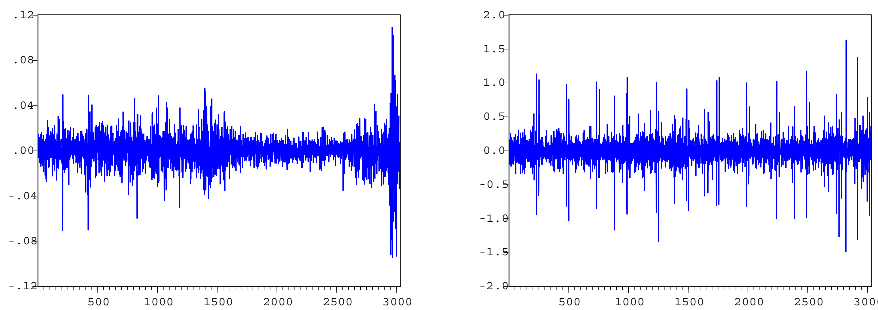
five percent significance level, we reject the non-causality hypothesis for all directions of causality (from returns to volatility and from volatility to returns) and all values of c. Contrary to the linear causality tests, we now confirm that both nonlinear leverage and volatility feedback effects can explain the asymmetric relationship between returns and volatility.

5.2 Application 2: Causality between returns and volume

The relationship between returns and volume has been subject to extensive theoretical and empirical research. Morgan (1976), Epps and Epps (1976), Westerfield (1977), Rogalski (1978), and Karpoff (1987) using daily or monthly data find a positive correlation between volume and returns (absolute returns). Gallant, Rossi, and Tauchen (1992) considering a semiparametric model for conditional joint density of market price changes and volume conclude that large price movements are followed by high volume. Hiemstra and Jones (1994) use non-linear Granger causality test proposed by Baek and Brock (1992) to examine the non-linear causal relation between volume and return and find that there is a positive bi-directional relation between them. However, Baek and Brock (1992)'s test assumes that the data for each individual variable is *i.i.d.* More recently, Gervais, Kaniel, and Mingelgrin (2001) show that periods of extremely high volume tend to be followed by positive excess returns, whereas periods of extremely low volume tend to be followed by negative excess returns. In this application, we reexamine the relationship between returns and volume using daily data on S&P 500 Index. First we test for linear causality and than we use our nonparametric tests to check whether there is nonlinear relationships between these two variables.

5.2.1 Data description

The data set comes from Yahoo Finance and consists of daily observations on the S&P 500 Index. The sample runs from January 1997 to January 2009 for a total of 3032 observations, see Figure 2 for the series in growth rates. We perform Augmented Dickey-Fuller tests (hereafter ADF-tests) for



(a) S&P 500 Index returns (b) S&P 500 Index volume growth

Figure 2: S&P 500 Index returns and volume growth rate. The sample covers the period from January 1997 to January 2009 for a total of 3032 observations.

nonstationarity of the logarithmic price and volume and their first differences. Using an ADF-test with only an intercept, the results show that all variables in logarithmic form are nonstationary. The test statistics for log price and log volume are -2.259 and -1.173 respectively and the corresponding critical value at 5% significance level is -2.863. However, their first differences are stationary. The test statistics for log price and log volume are -43.655 and -20.653, respectively. Using ADF-tests with both intercept and trend leads to the same conclusions. Based on the above stationarity tests, we model the first difference of logarithmic price and volume rather than their level. Consequently, the causality relations have to be interpreted in terms of the growth rates.

5.2.2 Causality tests

To test for linear causality between returns and volume we estimate a first order vector autoregressive model. This yields the following results [t-statistics are between brackets]

$$\begin{bmatrix} \hat{r}_t \\ \\ \widehat{\Delta \ln(V_t)} \end{bmatrix} = \begin{bmatrix} \underset{[0.20655]}{4.96\ 10^{-5}} \\ \\ \underset{[0.38206]}{0.001261} \end{bmatrix} + \begin{bmatrix} \underset{[-3.75188]}{-0.068044} & \underset{[0.40386]}{0.000503} \\ \\ \underset{[-5.36570]}{-1.338504} & \underset{[-18.8823]}{-0.323437} \end{bmatrix} \begin{bmatrix} r_{t-1} \\ \\ \Delta \ln(V_{t-1}) \end{bmatrix} \quad \begin{matrix} R^2 = 0.0047 \\ \\ R^2 = 0.1120 \end{matrix} \;. \tag{9}$$

Equation (9) shows that the causality from returns to volume is statistically significant at 5% significance level with t-statistic equal to -5.365 [For *p-values* see Table 5]. However, the feedback effect from volume to returns is statistically insignificant at the same significance level with t-statistic equal to 0.404. Considering different orders for vector autoregressive model leads to the same conclusion.

Since volume fails to have a linear impact on returns, next we examine the nonlinear relationships between these two variables by applying our nonparametric test. The *p-values* are presented in Table 5. The latter shows that, at 5% significance level, nonparametric test rejects clearly the null hypothesis of non-causality from returns to volume, which is in line with the conclusion from the linear test. Further, our nonparametric test also detects a non-linear feedback effect from volume to returns at 5% significance level.

Table 5: P-values for linear and nonlinear causality tests

Test statistic / H_0	returns to volume	volume to returns
LIN	0.000	0.654
BRT, $c = 1$	0.000	0.005
BRT, $c = 1.5$	0.000	0.045
BRT, $c = 2$	0.010	0.055

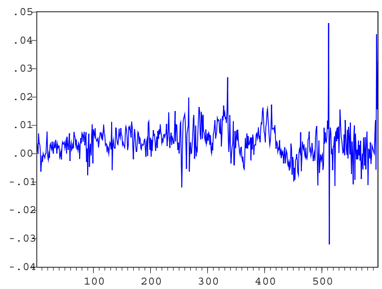
Linear and Nonlinear causality tests between returns (r) and volume (ln(V)). LIN and BRT correspond to linear test and our nonparametric test, respectively.

5.3 Application 3: Causality between money, income and prices

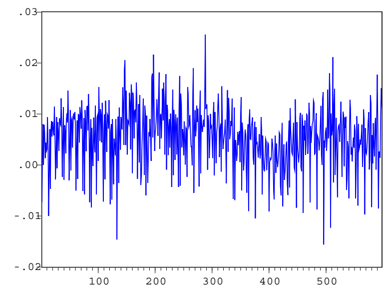
The relationships between money, income and prices have been the subject of a great deal of research over the last six decades. The approach commonly taken is based on the view that income and prices are related to the past and present values of money and vise versa. Sims (1972) shows using a reduced form model that money supply Granger causes income but that income does not Granger causes the money supply - thus lending support to the Monetarist viewpoint against the Keynesian viewpoint which claims that money does not play any role in changing income and prices. Sims' model has been reduced to a single equation relating income only to money, thereby ignoring the specific impacts of other variables. However, using a vector autoregressive model containing also interest rates and prices variables Sims (1980a) argues that while pre-war cycles do seem to support the Monetarist thesis, the post-war cycles are quite different. Specifically, he finds that in the post-war period, the interest rate accounted for most of the effect on output previously attributed to money. Bernanke and Blinder (1992) and Bernanke and Mihov (1998) also present evidence consistent with the view that the impact of monetary policy on the economy works through interest rates. In this application, we reanalyze the linear relationships between monetary and economic variables using U.S. data until November 2008 and we use our nonparametric test to check whether nonlinear relationships between these variables exist.

5.3.1 Data description

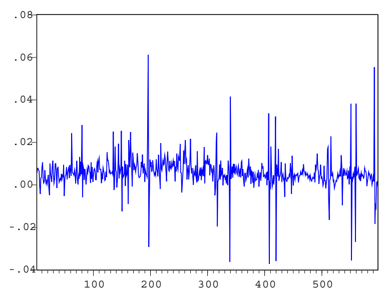
The data comes from the Federal Reserve Bank of St. Louis and consists of seasonally adjusted monthly observations on aggregates M1 and M2, disposable personal income (DPI), real disposable personal income (RDPI), industrial output (IP) and consumer price index (CPI). The sample runs from January 1959 to November 2008 for a total of 599 observations; see Figure 3 for the series in growth rates. Since all the variables in natural logarithms are nonstationary, we perform *ADF*-



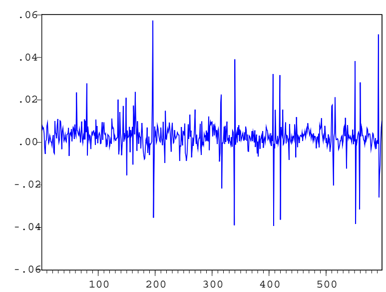
(a) M1 Money Stock



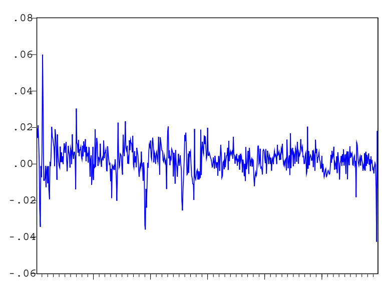
(b) M2 Money Stock



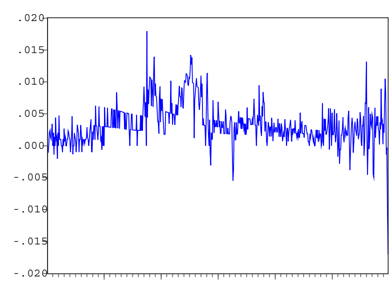
(c) Disposable Personal Income



(d) Real Disposable Personal Income



(e) Industrial Production Index



(f) Consumer Price Index

Figure 3: growth rates of the variables. The sample covers the period from January 1959 to November 2008 for a total of 599 observations.

test for nonstationarity of the growth rates of six variables. The results are presented in Table 6 and show that the growth rates of all variables are stationary except for the CPI. We perform a nonstationarity test for the second difference of variable CPI and find that the test statistic values are equal to -8.493 and -8.523 for the ADF-test with only an intercept and with both intercept and trend, respectively. The critical values are equal to -2.866 and -3.418, suggesting that the second difference of the CPI variable is stationary.

Table 6: Augmented Dickey-Fuller tests for growth rates

	With Intercept		With Intercept and Trend	
	test statistic	5% Critical Value	test statistic	5% Critical Value
M1	-4.530	-2.866	-4.460	-3.417
M2	-4.032	-2.866	-4.391	-3.418
DPI	-29.328	-2.866	-29.751	-3.417
RDPI	-17.782	-2.866	-18.000	-3.417
IP	-10.492	-2.866	-10.576	-3.417
CPI	-2.456	-2.866	-2.625	-3.417

Augmented Dickey-Fuller tests (ADF-test) for growth rates of monetary aggregates [M1 and M2] and economic aggregates [disposable personal income (DPI), real disposable personal income (RDPI), industrial output (IP) and consumer price index (CPI)].

5.3.2 Causality tests

To test linear Granger causality between money, income and prices, we consider the following models:

$$\Delta Y_t = \mu + \phi \Delta Y_{t-1} + \delta \Delta X_{t-1} + \varepsilon_t,$$

where $Y \neq X$ and $Y, X =$ M1, M2, DPI, RDPI, IP, CPI. More particularly, we examine the linear Granger causality from M1 (M2) to DPI, RDPI, IP, CPI, and vice versa. We say that ΔX does not Granger cause ΔY if $H_0 : \delta_1 = ... = \delta_p = 0$ is true. Similarly, we test linear Granger causality from ΔY to ΔX by considering a linear regression model where we regress ΔX on its own past and the past of ΔY. The results of linear Granger causality between money, income and prices are presented in Table 7 [see rows named LIN]. The Granger causality tests from M1 to DPI, RDPI, CPI are statistically insignificant at 5% level. The same conclusion is true for IP at 1% level. When M1 is replaced by M2, the causality directions from the economic variables DPI, RDPI, IP, CPI to M2 are statistically significant at 5% level. Further, the causality from M2 to DPI is statistically significant at the same level. The results do not change when we use the second rather than first difference of the CPI. As a conclusion, income and prices cause the monetary aggregates, while monetary policy measured by M1 and M2 has no impact on income and prices.

Since money fails to have a linear impact on income and prices, next we test for nonlinear relationships between these variables by applying our nonparametric test. The results are presented in Table 7. The latter shows that the nonlinear relationship from M1 to DPI is statistically significant at 5% level, even if the linear relationship does not exist at the same level. Same conclusion holds for causality from M2 to RDPI. We also find that the feedback from DPI to M2 and from RDPI to M2 are linearly and nonlinearly exist at 5% level.

Table 7: P-values for linear and nonlinear causality tests

H_0 :	DPI	RDPI	IP	CPI	DPI	RDPI	IP	CPI
	does not cause M1				does not cause M2			
LIN	0.116	0.219	0.759	0.116	0.007	0.009	0.899	0.406
BRT, $c = 1$	0.004	0.036	0.600	0.348	0.002	0.018	0.058	0.526
BRT, $c = 1.5$	0.036	0.160	0.302	0.248	0.028	0.033	0.006	0.426
BRT, $c = 2$	0.076	0.148	0.422	0.474	0.048	0.046	0.014	0.492
	M1 does not cause				M2 does not cause			
	DPI	RDPI	IP	CPI	DPI	RDPI	IP	CPI
LIN	0.297	0.870	0.027	0.462	0.058	0.204	0.881	0.561
BRT, $c = 1$	0.004	0.336	0.996	0.476	0.072	0.012	0.108	0.462
BRT, $c = 1.5$	0.048	0.674	0.948	0.368	0.096	0.012	0.114	0.676
BRT, $c = 2$	0.078	0.638	0.978	0.324	0.178	0.011	0.366	0.666

Linear and Nonlinear causality tests between monetary aggregates [M1 and M2] and economic aggregates [disposable personal income (DPI), real disposable personal income (RDPI), industrial output (IP) and consumer price index (CPI)]. LIN and BRT correspond to linear test and our nonparametric test, respectively.

6 Conclusion

A nonparametric copula-based test for conditional independence between two vector processes conditional on another one is proposed. The test statistic requires the estimation of the copula density functions. We consider a nonparametric estimator of the copula density based on Bernstein polynomials. The Bernstein copula estimator is always non-negative and does not suffer from boundary bias problem. Further, the proposed test is easy to implement because it does not involve a weighting function in the test statistic, and it can be applied in general settings since there is no restriction on the dimension of the data. To apply our test, only a bandwidth is needed for the nonparametric copula.

We show that the test statistic is asymptotically pivotal under the null hypothesis, we establish local power properties, and we motivate the validity of the bootstrap technique that we use in finite sample settings. A simulation study illustrates the good size and power properties of the test.

We consider several empirical applications to illustrate the usefulness of our nonparametric test. In these applications, we examine the Granger non-causality between many macroeconomic and financial variables. Contrary to the general findings in the literature, we provide evidence on two alternative mechanisms of nonlinear interaction between returns and volatilities: the nonlinear leverage effect and the nonlinear volatility feedback effect.

As a further extension of this paper, it would be interesting to investigate deeper the bandwidth selection for our nonparametric test, similar to the approach of Gao and Gijbels (2008) who test the equality between an unknown and a parametric mean function. While respecting the size, their bandwidth maximizes the power of the test. Another interesting direction would be to investigate analytically the small sample properties of our test along the lines of Fan and Linton (2003).

Appendix

In this appendix, we provide the proofs of the theoretical results developed in Section 3. We begin by studying the asymptotic distribution of the pseudo-statistic $H(\hat{c}, C)$ obtained by replacing $C_{XYZ,T}$ in $\hat{H}$ by C_{XYZ}. Thereafter, we show, see Lemma 4, that the distribution of $\hat{H}$ is given by the distribution of $H(\hat{c}, C)$. The main element in the proof of Theorem 1 and other propositions of Section 3 is the asymptotic normality of U-statistics. We use Theorem 1 of Tenreiro (1997) to prove our Theorem 1 and Proposition 1. For dependent data, the central limit theorem of the U-statistics is also investigated in Fan and Li (1999). To show the validity of the local smoothed bootstrap in Proposition 2, we use Theorem 1 of Hall (1984).

For simplicity of notation and to keep more space, in what follows we replace the notations $c_{XY}(u,v)$, $c_{XZ}(u,w)$, $c_{XYZ}(u,v,w)$, and $C_{XYZ}(u,v,w)$ by $c(u,v)$, $c(u,w)$, $c(u,v,w)$ and $C(u,v,w)$, respectively. We do the same with their estimates $\hat{c}_{XY}(u,v)$, $\hat{c}_{XZ}(u,w)$, $\hat{c}_{XYZ}(u,v,w)$ and $\hat{C}_{XYZ}(u,v,w)$. Without loss of generality and since

$$\hat{c}_{XYZ}(\bar{F}_{X,T}(x), \bar{F}_{X,T}(y), \bar{F}_{Z,T}(z)) = \hat{c}_{XYZ}(\bar{F}_X(x), \bar{F}_Y(y), \bar{F}_Z(z)) + O(T^{-1}),$$

in what follows we will work with

$$H(\hat{c}, C) = \frac{1}{T} \sum_{t=1}^{T} \left\{ 1 - \sqrt{\frac{\hat{c}_{XY}(U_t)\hat{c}_{XZ}(V_t)}{\hat{c}_{XYZ}(G_t)}} \right\}^2,$$

where $\bar{F}_X(.)$, $\bar{F}_Y(.)$, $\bar{F}_Z(.)$, $\bar{F}_{X,T}(.)$, $\bar{F}_{X,T}(.)$, $\bar{F}_{Z,T}(.)$ are defined in Section 2 and

$$G_t = (G_{t1}, ..., G_{td}) = (\bar{F}_X(X_t), \bar{F}_Y(Y_t), \bar{F}_Z(Z_t)),$$

$$U_t = (U_{t1}, ..., U_{t(d_1+d_2)}) = (\bar{F}_X(X_t), \bar{F}_Y(Y_t)),$$

$$V_t = (V_{t1}, ..., V_{t(d_1+d_3)}) = (\bar{F}_X(X_t), \bar{F}_Z(Z_t)).$$

We will show that the distribution of $\hat{H}$ is given by the distribution of $H(\hat{c}, C)$ which is stated by the following four lemmas. In the next lemma, we rewrite $H(\hat{c}, C)$ so that it becomes easier to work with.

Lemma 1 *Under assumptions **(A1.1)-(A1.3)** and H_0, we have*

$$H(\hat{c}, C) = \frac{1}{4}\int_{[0,1]^d}\left\{\frac{\hat{c}(u,v,w)}{c(u,v,w)} - \frac{c(u,v)}{c(u,v)} - \frac{\hat{c}(u,w)}{c(u,w)} + 1\right\}^2 dC(u,v,w) + O_P(||\hat{c}(u,v,w) - c(u,v,w)||_\infty^3),$$

where $||\hat{c}(u,v,w) - c(u,v,w)||_\infty = \sup_{(u,v,w)\in[0,1]^d} |\hat{c}(u,v,w) - c(u,v,w)|.$

Proof: Let's consider

$$\phi(\alpha) = \int \left(1 - \sqrt{\frac{\phi_1(\alpha)\phi_2(\alpha)}{\phi_3(\alpha)}}\right)^2 dC(u,v,w), \text{ for } \alpha \in [0,1]$$

and

$$\phi_1(\alpha) = c(u,v) + \alpha c^*(u,v),$$

$$\phi_2(\alpha) = c(u,w) + \alpha c^*(u,w),$$

$$\phi_3(\alpha) = c(u,v,w) + \alpha c^*(u,v,w),$$

where $c^*(u,v,w)$, $c^*(u,v)$ and $c^*(u,w)$ are functions in Γ_i for $i = 1$, 2 and 3 respectively and Γ_i is a set defined as

$$\Gamma_i = \left\{\gamma : [0,1]^{p_i} \to \mathbb{R},\ \gamma \text{ is bounded, } \int \gamma = 0 \text{ and } ||\gamma|| > b/2\right\}$$

with $p_i = d$, $d_1 + d_2$, $d_1 + d_3$, for $i = 1$, 2 and 3, respectively. Using Taylor's expansion, we have

$$\phi(\alpha) = \phi(0) + \alpha\phi'(0) + \frac{1}{2}\alpha^2\phi''(0) + \frac{1}{6}\alpha^3\phi'''(\alpha^*), \text{ for } \alpha^* \in [0,\alpha].$$

We can show that:

$$\phi'(\alpha) = \int \left(1 - \sqrt{\frac{\phi_3(\alpha)}{\phi_1(\alpha)\phi_2(\alpha)}}\right)\left\{\frac{c^*(u,v)\phi_2(\alpha) + c^*(u,w)\phi_1(\alpha)}{\phi_3(\alpha)} - \frac{c^*(u,v,w)\phi_1(\alpha)\phi_2(\alpha)}{\phi_3^2(\alpha)}\right\} dC(u,v,w),$$

$$\begin{aligned}\phi''(\alpha) \quad &= \quad \int \sqrt{\frac{\phi_1(\alpha)\phi_2(\alpha)}{\phi_3(\alpha)}}\left\{\frac{c^*(u,v)}{\phi_1(\alpha)} + \frac{c^*(u,w)}{\phi_2(\alpha)} - \frac{c^*(u,v,w)}{\phi_3(\alpha)}\right\}^2 dC(u,v,w) \\ &+ \left(1 - \sqrt{\frac{\phi_3(\alpha)}{\phi_1(\alpha)\phi_2(\alpha)}}\right)\frac{d}{d\alpha}\left\{\frac{c^*(u,v)\phi_2(\alpha) + c^*(u,w)\phi_1(\alpha)}{\phi_3(\alpha)} - \frac{c^*(u,v,w)\phi_1(\alpha)\phi_2(\alpha)}{\phi_3^2(\alpha)}\right\} dC(u,v,w)\end{aligned}$$

and

$$\phi'''(\alpha) = O(||c^*(u,v,w)||_\infty^3 + ||c^*(u,v)||_\infty^3 + ||c^*(u,w)||_\infty^3).$$

Under H_0, we have $\phi(0) = \phi'(0) = 0$ and

$$\phi''(0) = \int \left\{ \frac{c^*(u,v)}{c(u,v)} + \frac{c^*(u,w)}{c(u,w)} - \frac{c^*(u,v,w)}{c(u,v,w)} + 1 \right\}^2 dC(u,v,w).$$

Next, we consider $\alpha = 1$, $c^*(u,v,w) = \hat{c}(u,v,w) - c(u,v,w)$, $c^*(u,v) = \hat{c}(u,v) - c(u,v)$, and $c^*(u,w) = \hat{c}(u,w) - c(u,w)$. Using the results of Bouezmarni, Rombouts, and Taamouti (2009), we get

$$||\hat{c}(u,v,w) - c(u,v,w)||_\infty = O_p(T^{-1/2}k^{d/4}\ln^\theta(T) + k^{-1}) = o_p(1)$$

and for a positive constant θ, we can show that $c^*(u,v,w)$, $c^*(u,v)$ and $c^*(u,w)$ are in Γ_i for $i = 1$, 2 and 3, respectively. Since the term $||c^*(u,v,w)||_\infty$ dominates the terms $||c^*(u,v)||_\infty$ and $||c^*(u,w)||_\infty$, this concludes the proof of the lemma.

■

Now, let us take $g = (u,v,w)$ and denote by

$$\hat{c}(u,v,w) = \tfrac{1}{T}\textstyle\sum_{t=1}^{T} K^1(g,G_t),$$

$$\hat{c}(u,v) = \tfrac{1}{T}\textstyle\sum_{t=1}^{T} K^2(g,G_t),$$

$$\hat{c}(u,w) = \tfrac{1}{T}\textstyle\sum_{t=1}^{T} K^3(g,G_t),$$

$$\tilde{R}(g,m) = \textstyle\sum_{j=1}^{4}[R_j(g,m) - \mathbb{E}(R_j(g,m))],$$

with

$$R(g,m) = \frac{K^1(g,m)}{c(g)} - \frac{K^2(g,m)}{c(u,v)} - \frac{K^3(g,m)}{c(u,w)} + 1 = \sum_{j=1}^{4} R_j(g,m), \text{ for } m \in [0,1]^d$$

where

$$R_1(g,m) = \frac{K^1(g,m)}{c(g)},\ R_2(g,m) = -\frac{K^2(g,m)}{c(u,v)},\ R_3(g,m) = -\frac{K^3(g,m)}{c(u,w)},\ R_4(g,m) = K^4(g,m) = 1 \tag{10}$$

and

$$K^1(g,G_t) = K_k(g,G_t),\ K^2(g,G_t) = K_k((u,v),U_t),\ K^3(g,G_t) = K_k((u,w),V_t).$$

Further, consider

$$I_T = \int_{[0,1]^d} \left\{ \frac{\hat{c}(g)}{c(g)} - \frac{\hat{c}(u,v)}{c(u,v)} - \frac{\hat{c}(u,w)}{c(u,w)} + 1 \right\}^2 dC(g) = \int r_T^2(g)dC(g).$$

We show that

$$
\begin{aligned}
I_T - \mathbb{E}(I_T) &= 2\int (r_T(g) - \mathbb{E}(r_T(g)))\mathbb{E}(r_T(g))dC(g) + \int \left([r_T(g) - \mathbb{E}(r_T(g))]^2 - \mathbb{E}[r_T(g) - \mathbb{E}(r_T(g))]^2 \right) dC(g) \\
&= 2T^{-1/2}k^{-1}\left\{ T^{-1/2}\sum_{t=1}^{T} S_T(G_t) \right\} + 2T^{-1}k^{d/2}\left\{ T^{-1}\sum_{1\le t<s\le T} [H_T(G_t, G_s) - \mathbb{E}(H_T(G_t, G_j))] \right\} \\
&\quad + T^{-1}k^{d/2}\left\{ T^{-1}\sum_{1\le t\le T} [H_T(G_t, G_t) - \mathbb{E}(H_T(G_t, G_t))] \right\} \\
&= 2T^{-1/2}k^{-1}I_1 + 2T^{-1}k^{d/2}I_2 + T^{-1}k^{d/2}I_3,
\end{aligned}
$$

where

$$I_1 = T^{-1/2}\textstyle\sum_{t=1}^{T} S_T(G_t),$$

$$I_2 = T^{-1}\textstyle\sum_{1\le t<s\le T}[H_T(G_t, G_s) - \mathbb{E}(H_T(G_t, G_j))], \tag{11}$$

$$I_3 = T^{-1}\textstyle\sum_{1\le t\le T}[H_T(G_t, G_t) - \mathbb{E}(H_T(G_t, G_t))],$$

and

$$S_T(m) = k\textstyle\int \tilde{R}(g, m)\mathbb{E}(r(g))dC(g), \text{ for } m \in [0,1]^d$$

$$H_T(m_1, m_2) = k^{-d/2}\textstyle\int \tilde{R}(g, m_1)\tilde{R}(g, m_2)dC(g), \text{ for } m_1, m_2 \in [0,1]^d.$$

Next, we show that $I_3 = O_p(T^{-1/2}k^{d/4})$ and we apply Theorem 1 of Tenreiro (1997) for the other terms I_1 and I_2. However, observe that under Assumption (**A1.3)** on the bandwidth parameter, the term $2T^{-1/2}k^{-1}I_1$ is negligible. Hence, the asymptotic distribution of $T\,k^{-d/2}(I_T - \mathbb{E}(I_T))$ is the same as the asymptotic distribution of I_2.

In what follows, we denote by $\sum\limits_{\upsilon} = \sum\limits_{\upsilon_1=0}^{k-1} \ldots \sum\limits_{\upsilon_d=0}^{k-1}$. To show that the term I_3 defined in (11) is negligible, we first compute the variance of $H_T(G_t, G_t)$. By observing that the term R_1, defined in (10), is the dominant term among $R_2(.)$, $R_3(.)$ and $R_4(.)$, we have

$$
\begin{aligned}
Var(H_T(G_t, G_t)) &= k^{-d}\,Var\left(\int \frac{K^1(g, G_t)K^1(g, G_t)}{c(g)}\, dg \right) \\
&= k^{-d}\,Var\left(\int \sum_{\upsilon} k^{2d} A_{G_t, \upsilon_1, .., \upsilon_d} \frac{\prod_{j=1}^{d} p_{\upsilon_j}^2(g_j)}{c(g)}\, dg \right) \\
&= \sum_{\upsilon}\left(k^{3d}\int (p_\upsilon - p_\upsilon^2)\frac{\prod_{j=1}^{d} p_{\upsilon_j}^4(g_j)}{c^2(g)}\, dg \right),
\end{aligned}
$$

where

$$p_\upsilon = \int_{\frac{\upsilon_d}{k}}^{\frac{\upsilon_d+1}{k}} ... \int_{\frac{\upsilon_1}{k}}^{\frac{\upsilon_1+1}{k}} c(u)du \tag{12}$$
$$= \frac{c(\frac{\upsilon_1}{k}, ..., \frac{\upsilon_d}{k})}{k^d} + O(k^{d+1}), \quad \text{from Sancetta and Satchell (2004).}$$

Consequently

$$Var(H_T(G_t, G_t))$$

$$\leq \int \left(\frac{k^{2d}}{c^2(g)} \sum_\upsilon c(\frac{\upsilon_1}{k}, ..., \frac{\upsilon_d}{k}) \prod_{j=1}^d p^4_{\upsilon_j}(g_j)\, dg \right)$$
$$\leq \frac{1}{\inf_g \{c_{XYZ}(g)\}} \int \left(\frac{k^d}{c(g)} \sum_\upsilon c(\frac{\upsilon_1}{k}, ..., \frac{\upsilon_d}{k}) \prod_{j=1}^d p^2_{\upsilon_j}(g_j)\, dg \right)^2$$
$$= \frac{k^d}{\inf_g \{c_{XYZ}(g)\}} \int \frac{1}{4\pi g(1-g)}\, dg, \quad \text{from Bouezmarni, Rombouts, and Taamouti (2009)}$$
$$= O(k^d).$$

Hence

$$I_3 = O_p(T^{-1/2}\, k^{d/2}).$$

■

The next lemma establishes the independence between the two terms I_1 and I_2 defined in (11) and their asymptotic normality. Further, under condition **(A1.3),** we show that $T^{-1/2}k^{-1}I_1$ is negligible. The following notations will be used to prove the lemma**.** For $p > 0$ and $\{\bar{G}_t, t \geq 0\}$ $i.i.d$ sequence, where $\bar{G}_0$ is an independent copy of G_0, we define

$$u_T(p) = \max\left\{ \max_{1\leq t\leq T} ||H_T(G_t, G_0)||_p, ||H_T(G_0, \bar{G}_0)||_p \right\}, \tag{13}$$

$$v_T(p) = \max\left\{ \max_{1\leq t\leq T} ||\Psi_{T,0}(G_t, G_0)||_p, ||\Psi_{T,0}(G_0, \bar{G}_0)||_p \right\}, \tag{14}$$

$$w_T(p) = \{||\Psi_{T,0}(G_0, G_0)||_p\}, \tag{15}$$

$$z_T(p) = \max_{0\leq t\leq T} \max_{1\leq j\leq T} \max\left\{ ||\Psi_{T,j}(G_t, G_0)||_p, ||\Psi_{T,j}(G_0, G_t)||_p, ||\Psi_{T,j}(G_0, \bar{G}_0)||_p \right\}, \tag{16}$$

where $\Psi_{T,j}(u, v) \equiv \mathbb{E}\left[H_T(G_t, u)H_T(G_0, v)\right]$ and $||.||_p = (\mathbb{E}|\,.\,|^p)^{1/p}$.

Lemma 2 *Under assumptions **(A1.1)-(A1.3)** and H_0, we have I_1 and I_2 are independents and*

$$I_1 \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

$$\sqrt{2}\,(\pi/4)^{d/2}\, I_2 \xrightarrow{d} N(0,1),$$

where $\sigma^2 = Var(\zeta(G_0)) + 2\sum_{t=1}^{\infty} Cov(\zeta(G_t), \zeta(G_0))$ and ζ is defined below.

Proof: To establish Lemma 2, we follow Theorem 1 in Tenreiro (1997). Recall that $I_1 = T^{-1/2}\sum_{t=1}^{T} S_T(G_t)$, where $S_T(G_t) = k\int \tilde{R}(g, G_t)\mathbb{E}(r(g))dC(g)$. By construction $\mathbb{E}(S_T(G_t)) = 0$ and by the boundedness of the copula density, $\sup_T \sup_m |S_T(m)| < \infty$. We have,

$$\begin{aligned}
\mathbb{E}(r(g)) &= \mathbb{E}\left(\frac{K^1(g, G_t)}{c(g)} - \frac{K^2(g, U_t)}{c(u,v)} - \frac{K^3(g, V_t)}{c(u,w)} + 1\right) \\
&= k^{-1}\gamma(g) + o(k^{-1}) \quad \text{from Bouezmarni, Rombouts, and Taamouti (2009),}
\end{aligned}$$

where $\gamma(g) = \frac{\gamma^*(c(g))}{c(g)} - \frac{\gamma^*(c(u,v))}{c(u,v)} - \frac{\gamma^*(c(u,w))}{c(u,w)}$ and $\gamma^*(g) = \frac{1}{2}\sum_{j=1}^{d}\left\{\frac{dc(g)}{dg_j}(1-2g_j) + \frac{d^2c(g)}{dg_j^2}g_j(1-g_j)\right\}$. Observe that,

$$\lim_T \mathbb{E}(S_T(G_t)S_T(G_0)) = Cov(\zeta(G_t), \zeta(G_0))$$

where $\zeta(m) = \int \gamma(g)c(g)\tilde{R}_j(g,m)dg$, for $m \in [0,1]^d$. Hence, under condition **(A1.3),** we have $2T^{-1/2}k^{-1}I_1 = o(T^{-1}k^{d/2})$, this concludes the proof of the lemma.

■

Now, it remains to show that there exist positive constants δ_0, δ_1, γ_1 and $\gamma_0 < 1/2$ such that **(1)** $u_T(4+\delta_0) = O(T^{\gamma_0})$; **(2)** $v_T(2) = o(1)$; **(3)** $w_T(2+\delta_0/2) = o(T^{1/2})$; **(4)** $z_T(2)T^{\gamma_1} = O(1)$; and thereafter we show that $\mathbb{E}\left[H_T(G_0,\bar{G_0})\right]^2 = \left(\frac{\pi}{2}\right)^d$. First, we show that

$$\begin{aligned}
k^d\mathbb{E}\left[H_T(G_0,\bar{G_0})\right]^2 &= \mathbb{E}\left\{\int \tilde{R}(g,G_0)\tilde{R}(g,\bar{G_0})c(g)dg\right\}^2 \qquad (17)\\
&\approx \mathbb{E}\left\{\int \frac{K^1(g,G_0)K^1(g,\bar{G_0})}{c(g)}dg\right\}^2, \quad \text{the other terms are negligibles}\\
&= \mathbb{E}\left\{\int \frac{K^1(g,G_0)K^1(g',G_0)K^1(g,\bar{G_0})K^1(g',\bar{G_0})}{c(g)c(g')}dgdg'\right\}\\
&= \mathbb{E}\left\{\int \frac{1}{c(g)c(g')}\left(k^{2d}\sum_{\upsilon} A_{G_0,\upsilon}\prod_{j=1}^{d} p_{\upsilon_j}(g_j)p_{\upsilon_j}(g'_j)\right)\right.\\
&\quad \left.\times\left(k^{2d}\sum_{\upsilon'} A_{\bar{G_0},\upsilon'}\prod_{j=1}^{d} p_{\upsilon'_j}(g_j)p_{\upsilon'_j}(g'_j)\right)dgdg'\right\}\\
&= \int \frac{1}{c(g)c(g')}\left(k^{2d}\sum_{\upsilon} p_{\upsilon}\prod_{j=1}^{d} p_{\upsilon_j}(g_j)p_{\upsilon_j}(g'_j)\right)\left(k^{2d}\sum_{\upsilon'} p_{\upsilon'}\prod_{j=1}^{d} p_{\upsilon'_j}(g_j)p_{\upsilon'_j}(g'_j)\right)dgdg'\\
&\approx \int \frac{1}{c(g)c(g')}\left(k^{d}\sum_{\upsilon} c(\upsilon_1/k,...,\upsilon_1/k)\prod_{j=1}^{d} p_{\upsilon_j}(g_j)p_{\upsilon_j}(g'_j)\right)\\
&\quad \times\left(k^{d}\sum_{\upsilon'} c(\upsilon'_1/k,...,\upsilon'_1/k)\prod_{j=1}^{d} p_{\upsilon'_j}(g_j)p_{\upsilon'_j}(g'_j)\right)dgdg',
\end{aligned}$$

where p_υ is defined in (12). Now, let's denote by $U = \{\upsilon,\ \text{for all } j,\ |\frac{\upsilon_j}{k} - g_j| < k^{-\delta} \text{ and } |\frac{\upsilon_j}{k} - g'_j| < k^{-\delta}\}$, for some positive constant δ and observe that:

$$\sum_{\upsilon} k^d c(\upsilon_1/k,...,\upsilon_1/k)\prod_{j=1}^{d} p_{\upsilon_j}(g_j)p_{\upsilon_j}(g'_j) = \sum_{\upsilon\in U}(.) + \sum_{\upsilon\in U^c}(.). \qquad (18)$$

Since $\upsilon \in U$ and for large k, we have $\frac{\upsilon_j}{k} \approx g_j \approx g'_j$. Hence, the first sum of Equation (18) can be written as follows

$$\begin{aligned}
\sum_{\upsilon\in U} k^d c(\upsilon_1/k,...,\upsilon_1/k)\prod_{j=1}^{d} p_{\upsilon_j}(g_j)p_{\upsilon_j}(g'_j) &\approx k^d c(g)\sum_{\upsilon\in U}\prod_{j=1}^{d} p^2_{\upsilon_j}(g_j)\\
&= k^d c(g)\prod_{j=1}^{d}\sum_{|\frac{\upsilon_j}{k}-g_j|<k^{-\delta}} p^2_{\upsilon_j}(g_j)\\
&\approx k^{d/2}c(g)\prod_{j=1}^{d}(4\pi g_j(1-g_j))^{-1/2}
\end{aligned}$$

from Bouezmarni, Rombouts, and Taamouti (2009).

Let's denote by $J=\left\{j,\ |\frac{\upsilon_j}{k}-g_j|>k^{-\delta}\right\}$. Without loss of generality, suppose that $|\frac{\upsilon_j}{k}-g'_j|<k^{-\delta}$, for all j in J, and that J contains $k_0>0$ elements. Thus, for the second sum of Equation (18) and using the results of Bouezmarni, Rombouts, and Taamouti (2009), we have

$$\begin{aligned}
\sum_{\upsilon\in U^c} k^d c(\upsilon_1/k,...,\upsilon_1/k)\prod_{j=1}^{d} p_{\upsilon_j}(g_j)p_{\upsilon_j}(g'_j) &\leq k^d\sup_g|c(g)|\left\{\prod_{j\in J}\left(\sum_{|\frac{\upsilon_j}{k}-g_j|>k^{-\delta}} p_{\upsilon_j}(g_j)\right)\right\}\\
&\quad\times\left\{\prod_{j\in J^c}\left(\sum_{|\frac{\upsilon_j}{k}-g_j|<k^{-\delta}} p^2_{\upsilon_j}(g_j)\right)\right\}\\
&\leq k^d\sup_g|c(g)|\left\{Ck^{-2k_0}\right\}\left\{C'k^{-(d-k_0)/2}\right\}\\
&= O(k^{d/2-3k_0/2})=o(k^{d/2}).
\end{aligned}$$

Consequently

$$\sum_{\upsilon} k^d c(\upsilon_1/k,...,\upsilon_1/k)\prod_{j=1}^{d} p_{\upsilon_j}(g_j)p_{\upsilon_j}(g'_j) = k^{d/2}c(g)\prod_{j=1}^{d}(4\pi g_j(1-g_j))^{-1/2}+o(k^{d/2}).$$

Hence

$$\begin{aligned}
\lim_T\mathbb{E}\left[H_T(G_0,\bar{G}_0)\right]^2 &= \lim_T k^{-d}\int\frac{1}{c(g)c(g')}\left(k^{d/2}c(g)\prod_{j=1}^{d}(4\pi g_j(1-g_j))^{-1/2}+o(k^{d/2})\right)\\
&\quad\times\left(k^{d/2}c(g')\prod_{j=1}^{d}(4\pi g'_j(1-g'_j))^{-1/2}+o(k^{d/2})\right)dgdg'\\
&= (\pi/4)^d.
\end{aligned}$$

■

Now, let's calculate the term $u_T(p)$ in (13). Using the triangle inequality and since $K^1(g, G_0)K^1(g, G_t)$ is the dominant term [$K^1(g, G_t)$ is the dominant kernel], we have

$$\begin{aligned} ||H_T(G_0, G_t)||_p &\le k^{-d/2} \sum_{j=1}^{4} \sum_{j'=1}^{4} \left|\left| \int \frac{K^j(g, G_0)K^{j'}(g, G_t)}{c_j(g)c_{j'}(g)} c(g) dg \right|\right|_p \\ &= k^{-d/2} \left|\left| \int \frac{K^1(g, G_0)K^1(g, G_t)}{c(g)} dg \right|\right|_p + o(k^{-d/2}), \end{aligned}$$

where $K^4(g, G_t) = 1$. Since p_υ, $p_{\upsilon'} \le 1$ and $\sum_\upsilon \prod_{j=1}^d p_{\upsilon'_j}(g_j) = \sum_{\upsilon'} \prod_{j=1}^d p_{\upsilon'_j}(g_j) = 1$, by the triangular inequality of L_p we have

$$\begin{aligned} \left|\left| \int \frac{K^1(g, G_0)K^1(g, G_t)}{c(g)} dg \right|\right|_p &\le k^{2d} \sum_{\upsilon} \sum_{\upsilon'} \left|\left| \int \frac{\mathbf{1}_{\{G_0 \in B_\upsilon, G_t \in B_{\upsilon'}\}} \prod_{j=1}^d p_{\upsilon_j}(g_j) p_{\upsilon'_j}(g_j)}{c(g)} dg \right|\right|_p \\ &\le k^{2d} \sum_{\upsilon} \sum_{\upsilon'} \left(\left| \int \frac{\prod_{j=1}^d p_{\upsilon_j}(g_j) p_{\upsilon'_j}(g_j)}{c(g)} dg \right|^p p_\upsilon p_{\upsilon'} \right)^{1/p} \\ &= k^{2d} \sum_{\upsilon} \sum_{\upsilon'} p_\upsilon^{1/p} p_{\upsilon'}^{1/p} \int \frac{\prod_{j=1}^d p_{\upsilon_j}(g_j) p_{\upsilon'_j}(g_j)}{c(g)} dg \\ &= O(k^{2d}) \end{aligned}$$

Therefore, we have $4||H_T(G_0, G_t)||_p = O(k^{\frac{3d}{2}})$. Similarly, we can show that $||H_T(G_0, \bar{G}_0)||_p = O(k^{\frac{3d}{2}})$. Hence, $u_T(p) = O(k^{\frac{3d}{2}})$.

To compute the term $v_T(p)$ in (14) we need to calculate $||H_T(G_0, G_0)H_T(G_0, \bar{G}_0)||_p$. Using the fact that $K^1(g, G)$ is the dominant kernel, we have

$$\begin{aligned} H_T(G_0, G_0)H_T(G_0, \bar{G}_0) &= k^{-d} \int\int \frac{K_k^2(g, G_0)K_k(g', G_0)K_k(g', \bar{G}_0)}{c(g)c(g')} dg\, dg') \\ &= k^{-d} \int\int \frac{1}{c(g')c(g)} \left(k^d \sum_{\upsilon} \mathbf{1}_{\{G_0 \in B_\upsilon\}} \prod_{j=1}^{d} p_{\upsilon_j}(g_j) \right)^2 \\ &\quad \left(k^d \sum_{\upsilon} \mathbf{1}_{\{G_0 \in B_\upsilon\}} \prod_{j=1}^{d} p_{\upsilon_j}(g'_j) \right) \left(k^d \sum_{\upsilon'} \mathbf{1}_{\{\bar{G}_0 \in B_\upsilon\}} \prod_{j=1}^{d} p_{\upsilon_j}(g'_j) \right) dg\, dg' \\ &= k^{-d} \sum_{\upsilon} \sum_{\upsilon'} \int\int \frac{1}{c(g')c(g)} \left(k^d \mathbf{1}_{\{G_0 \in B_\upsilon\}} \prod_{j=1}^{d} p_{\upsilon_j}(g_j) \right)^2 \\ &\quad \left(k^d \mathbf{1}_{\{G_0 \in B_\upsilon\}} \prod_{j=1}^{d} p_{\upsilon_j}(g'_j) \right) \left(k^d \mathbf{1}_{\{\bar{G}_0 \in B_\upsilon\}} \prod_{j=1}^{d} p_{\upsilon'_j}(g'_j) \right) dg\, dg'. \end{aligned}$$

By the triangular inequality of L_p, we have

$$||H_T(G_0,G_0)H_T(G_0,\bar{G}_0)||_p \leq k^{-d}\sum_{\upsilon}\sum_{\upsilon'}\left|\left|B_{\upsilon,\upsilon'}\right|\right|_p$$

where

$$\begin{aligned}B_{\upsilon,\upsilon'} &= \int\int \frac{1}{c(g')c(g)}\left(k^d\mathbf{1}_{\{G_0\in B_\upsilon\}}\prod_{j=1}^d p_{\upsilon_j}(g_j)\right)^2\left(k^d\mathbf{1}_{\{G_0\in B_\upsilon\}}\prod_{j=1}^d p_{\upsilon_j}(g'_j)\right)\\ &\times\left(k^d\mathbf{1}_{\{\bar{G}_0\in B_{\upsilon'}\}}\prod_{j=1}^d p_{\upsilon'_j}(g'_j)\right)dg\,dg'.\end{aligned}$$

Consequently

$$\begin{aligned}\sum_{\upsilon}\sum_{\upsilon'}||B_{\upsilon,\upsilon'}||_p &= \sum_{\upsilon}\sum_{\upsilon'}\left\{\int\int\frac{1}{c(g')c(g)}\left(k^d\prod_{j=1}^d p_{\upsilon_j}(g_j)\right)^2\left(k^d\prod_{j=1}^d p_{\upsilon_j}(g'_j)\right)\right.\\ &\left.\times\left(k^d\prod_{j=1}^d p_{\upsilon'_j}(g'_j)\right)dg\,dg'\right\}p_\upsilon^{1/p}p_{\upsilon'}^{1/p}\\ &= k^{-2d/p}\sum_{\upsilon}\sum_{\upsilon'}\left\{\int\int\frac{1}{c(g')c(g)}\left(k^d\prod_{j=1}^d p_{\upsilon_j}(g_j)\right)^2\left(k^d\prod_{j=1}^d p_{\upsilon_j}(g'_j)\right)\right.\\ &\left.\times\left(k^d\prod_{j=1}^d p_{\upsilon'_j}(g'_j)\right)dg\,dg'\right\}c^{1/p}(\upsilon_1/k,...,\upsilon_d/k)c^{1/p}(\upsilon'_1/k,...,\upsilon'_d/k).\end{aligned}$$

Thereafter, $||B_{\upsilon,\upsilon'}||_p = O(k^{-d(1/p-1/2)})$ and $||H_T(G_0,G_0)H_T(G_0,\bar{G}_0)||_p = O(k^{-d(1/p+1/2)})$. Hence, $v_T(p) = O(k^{-d(1/p+1/2)})$. Similarly we can show that for the terms $w_T(p)$ and $z_T(p)$ in (15) and (16) respectively we have: $w_T(p)\leq Constant$ and $z_T(p) = O(k^d)$.

■

The following lemma provides the bias terms of the test statistic.

Lemma 3 *Under assumptions **(A1.1)-(A1.3)** and H_0, we have*

$$Tk^{-d/2}\mathbb{E}(I_T) = \frac{3}{2}\pi^{d-1/2} + o(1).$$

Proof: Observe that

$$\mathbb{E}(I_T) = \mathbb{E}\int\left(r_T(g)-\mathbb{E}(r_T(g))\right)^2c(g)\,dg + \int\mathbb{E}(r_T(g))^2c(g)\,dg.$$

Under Assumption (**A1.3)** on the bandwidth parameter and using the bias of the Bernstein density copula estimator [see Bouezmarni, Rombouts, and Taamouti (2009)], we have

$$T\, k^{-d/2}\mathbb{E}(r_T(g))^2 = O(T\, k^{-d/2-2}) = o(1).$$

Since $r_T(g) - \mathbb{E}(r_T(g)) = \frac{1}{T}\sum_{t=1}^{T}\tilde{R}(g, G_t)$**,** we have

$$\begin{aligned}
\mathbb{E}\int (r_T(g) - \mathbb{E}(r_T(g)))^2 c(g)dg &= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}\int \tilde{R}^2(g, G_t)c(g)dg \\
&\quad + \frac{2}{T^2}\sum_{t<s}\mathbb{E}\int \tilde{R}(g, G_t)\tilde{R}(g, G_s)c(g)dg \\
&= T^{-1}k^{d/2}\left(\mathbb{E}(H_T(G_0, G_0)) + \frac{2}{T}\sum_{t<s}\mathbb{E}(H_T(G_t, G_s))\right).
\end{aligned}$$

We first show that

$$\begin{aligned}
\mathbb{E}(H_T(G_0, G_0)) &= k^{-d/2}\sum_{1\le j,j'\le 4}\mathbb{E}\left(\int \frac{K^j(g, G_0)K^{j'}(g, G_0)}{c_j(g)c_{j'}(g)}c(g)\, dg\right) + o(k^{-d/2}) \\
&= k^{-d/2}\sum_{j=1}^{10} D_j
\end{aligned}$$

where $c_1(g) = c(g)$, $c_2(g) = c(u, v)$, $c_3(g) = c(u, w)$, $c_4(g) = 1$ and

$$D_j = \begin{cases}
\mathbb{E}\int \left(\frac{K^j(g,G_0)}{c_j(g)}\right)^2 c(g)\, dg, & \text{for } j = 1,\ 2,\ 3,\ 4 \\
2\mathbb{E}(\int \frac{K^1(g,G_0)K^{j-3}(g,G_0)}{c_{j-3}(g)}\, dg), & \text{for } j = 5,\ 6,\ 7 \\
2\mathbb{E}(\int \frac{K^2(g,G_0)K^{j-5}(g,G_0)}{c_2(g)c_{j-3}(g)}c(g)\, dg), & \text{for } j = 8,\ 9 \\
2\mathbb{E}(\int \frac{K^3(g,G_0)K^4(g,G_0)}{c_3(g)c_4(g)}c(g)\, dg), & \text{for } j = 10.
\end{cases}$$

We can show that $D_4 = 1$, $D_7 = 2$, $D_9 = 2 + o(k^{-(d_1+d_2)})$ and $D_{10} = 2 + o(k^{-(d_1+d_3)})$. For the remainder, we use the properties of Bernoulli random variable and the variance of the Bernstein density estimator. Observe that $v \neq v'$ implies that $\mathbf{1}_{\{G_0 \in A_v\}} \times \mathbf{1}_{\{G_0 \in A_{v'}\}} = 0$. For D_1, we have

$$
\begin{aligned}
D_1 &= \mathbb{E} \int \frac{K^1(g, G_0) K^1(g, G_0)}{c_1^2(g)} c(g)\, dg \\
&= \mathbb{E} \int \frac{k^{2d} \sum_v \mathbf{1}_{\{G_0 \in A_v\}} \prod_{j=1}^d p_{v_j}^2(g_j)}{c(g)}\, dg \\
&= k^{d/2} \int \frac{c(g)}{c(g)\sqrt{4\pi \prod_{j=1}^d g_j(1-g_j)}}\, dg + o(k^{d/2}) \\
&= k^{d/2} (4\pi)^{d/2} \pi^d + o(k^{d/2}) \\
&= 2^{-d} \pi^{d/2} k^{d/2} + o(k^{d/2}).
\end{aligned}
$$

Using similar arguments, we can show that

$$
D_2 = k^{(d_1+d_2)/2} \int \frac{\prod_{j=1}^{d_1+d_2} (4\pi\, u_j(1-u_j))^{-1/2}}{c_2(g)} c(g)\, dg + o(k^{(d_1+d_2)/2}),
$$

and

$$
D_3 = k^{(d_1+d_3)/2} \int \frac{\prod_{j=1}^{d_1+d_3} (4\pi\, u_j(1-u_j))^{-1/2}}{c_3(g)} c(g)\, dg + o(k^{(d_1+d_3)/2}).
$$

For D_5, let's first denote by $A_{v'} = \left[\frac{v'_1}{k}, \frac{v'_1+1}{k}\right] \times ... \times \left[\frac{v'_{d_1+d_2}}{k}, \frac{v'_{d_1+d_2}+1}{k}\right]$ and $A^*_{v''} = \left[\frac{v''_1}{k}, \frac{v''_1+1}{k}\right] \times ... \times \left[\frac{v''_{d_3}}{k}, \frac{v''_{d_3}+1}{k}\right]$. Given $G_0 = (U_0, W_0)$ and using the variance and the bias of the Bernstein density copula estimator, we get

$$
\begin{aligned}
D_5 &= 2\mathbb{E} \int \frac{K^1(g, G_0) K^2(g, G_0)}{c_2(g)}\, dg \\
&= 2\mathbb{E} \int \frac{\left(k^d \sum_v \mathbf{1}_{\{G_0 \in B_v\}} \prod_{j=1}^d p_{v_j}(g_j)\right) \left(k^{d_1+d_2} \sum_{v'} \mathbf{1}_{\{U_0 \in A_{v'}\}} \prod_{j=1}^{d_1+d_2} p_{v'_j}(g_j)\right)}{c_2}\, dg \\
&= 2\mathbb{E} \int \frac{\left(k^{2(d_1+d_2)} \sum_{v'} \mathbf{1}_{\{U_0 \in A_{v'}\}} \prod_{j=1}^{d_1+d_2} p_{v'_j}^2(g_j)\right) \left(k^{d_3} \sum_{v''} \mathbf{1}_{\{W_0 \in A^*_{v''}\}} \prod_{j=1}^{d_3} p_{v''_j}(g_j)\right)}{c_2(g)}\, dg \\
&= 2k^{(d_1+d_2)/2} \int \prod_{j=1}^{d_1+d_2} (4\pi\, u_j(1-u_j))^{-1/2}\, du \int c(w) dw + o(k^{(d_1+d_2)/2}), \\
&= 2^{-(d_1+d_2-1)} \pi^{(d_1+d_2)/2} k^{(d_1+d_2)/2} + o(k^{(d_1+d_2)/2}).
\end{aligned}
$$

Similarly, we have

$$D_6 = 2^{-(d_1+d_3-1)}\pi^{(d_1+d_3)/2}k^{(d_1+d_3)/2} + o(k^{(d_1+d_3)/2}),$$

and

$$D_8 = 2k^{d_1/2}\int \frac{c(u)c(g)}{(4\pi)^{d_1/2}\prod_{j=1}^{d_1} u_j(1-u_j))}dg + o(k^{d_1/2}).$$

Hence

$$\mathbb{E}(H_T(G_0,G_0)) = k^{-d/2}\left[2^{-d}\pi^{d/2}k^{d/2} + k^{(d_1+d_2)/2}B_1 + k^{(d_1+d_3)/2}B_2 + k^{d_1/2}B_3 + 7\right] + o(1),$$

where $B1 = D_2 + D_5$, $B_2 = D_3 + D_5$ and $B_3 = D_8$. Now, we need to show that $\frac{2}{T}\sum_{t<s}\mathbb{E}(H_T(G_t,G_s)) = o(1)$. Denote by $n = [L\log T]$, where L is a large positive constant such that $T^4\beta_n^{\delta/(1+\delta)} = o(1)$ for some $\delta > 0$. On the one hand, from Lemma 1 of Yoshihara (1976) and by the fact that $u_T(1+\delta) \leq Ck^{d(1/2-1/p)}$, we obtain

$$\sum_{s-t>n}\mathbb{E}(H_T(G_t,G_s)) < CT^{-1}k^{d(1/2-1/p)}\beta_n^{\delta/(1+\delta)} = o(1).$$

On the other hand, using the fact that $u_T(1) = O(k^{-d/2})$, we get

$$\sum_{0<s-t<n}\mathbb{E}(H_T(G_t,G_s)) = O(nk^{-d/2}) = o(1),$$

and this concludes the proof of the lemma.

■

Lemma 4 *Under assumptions **(A1.1)-(A1.3)** and H_0, we have*

$$Tk^{-d/2}\left(H(\hat{c},\hat{C}) - H(\hat{c},C)\right) = o_p(1).$$

Proof: Lemma 4 can be deduced using similar arguments as in Ait-Sahalia, Bickel, and Stoke (2001) and Su and White (2008) and the properties of the Bernstein density copula estimator in Bouezmarni, Rombouts, and Taamouti (2009).

■

Proof of Proposition 1 This result can be obtained using a similar argument as in the proof of Theorem 1 and the Taylor expansion to calculate $H(c^{[T]}, C^{[T]})$ under $H_1(\alpha_T)$ and $H_1(\beta_T,\gamma_T)$.

■

Proof of Proposition 2 We obtain this result by using a similar argument as in the proof of Theorem 1, by replacing the terms, I_1, I_2 and I_3 in (11) by I_1^*, I_2^* and I_3^* obtained using the bootstrap data $\mathcal{G}^* = \{G_t^*\}_{t=1}^T$. Conditional on $\mathcal{G} = \{G_t\}_{t=1}^T$, $\{G_t^*\}$ forms a triangular array of independent random variables, hence $S_T(G_t)$ and $H_T(G_t,G_t)$ are independent. Further, conditional on $\mathcal{G}$ and using Theorem 1 of Hall (1984), we can show that I_2^* is asymptotically normal with mean zero and variance σ^2.

References

Ait-Sahalia, Y., P. Bickel, and T. Stoke (2001): "Goodness-of-fit Tests for Kernel Regression with an Application to Option Implied Volatilities," *Journal of Econometrics*, 105, 363–412.

Andersen, T., and T. Bollerslev (1998): "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2003): *Parametric and Non-Parametric Volatility Measurement*. In Handbook of Financial Econometrics, Vol. I, Ait-Sahalia Y, Hansen LP (eds). Elsevier-North Holland: Amsterdam., New York.

Baek, E., and W. Brock (1992): "A General Test for Non-Linear Granger Causality: Bivariate Model," Discussion paper, Iowa State University and University of Wisconsin, Madison, WI.

Barndorff-Nielsen, O., S. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2005): "A Central Limit Theorem for Realized Power and Bipower Variations of Continuous Semimartingales," Working Paper, Nuffield College, Oxford University; forthcoming in Yu Kabanov and Robert Liptser (eds.), From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev. New York: Springer-Verlag.

Barndorff-Nielsen, O., and N. Shephard (2002a): "Econometric Analysis of Realized Volatility and its use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, Series B*, 64, 253–280.

——— (2002b): "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*, 17, 457–478.

——— (2003): *Power and Bipower Variation with Stochastic and Jumps*. Manuscript, Oxford University.

Beran, R. (1977): "Minimum Hellinger Distance Estimates for Parametric Models," *The Annals of Statistics*, 5, 445–463.

Bernanke, B., and A. Blinder (1992): "The Federal Funds Rate and the Channels of Monetary Transmission," *American Economic Review*, 82, 901–921.

Bernanke, B., and I. Mihov (1998): "Measuring Monetary Policy," *The Quarterly Journal of Economics*, 113, 869–902.

Bierens, H., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

Black, F. (1976): "Studies of Stock Price Volatility Changes," *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics.*

Blum, J., J. Kiefer, and M. Rosenblatt (1961): "Distribution Free Tests of Independence based on the Sample Distribution Function," *Annals of Mathematical Statistics*, 32, 485–498.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T., J. Litvinova, and G. Tauchen (2006): "Leverage and Volatility Feedback Effects in High-Frequency Data," *Journal of Financial Econometrics*, 4(3), 353–384.

Bouezmarni, T., J. Rombouts, and A. Taamouti (2009): "Asymptotic Properties of the Bernstein Density Copula Estimator for alpha mixing Data," *Forthcoming in Journal of Multivariate Analysis.*

Campbell, J., and L. Hentschel (1992): "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics*, 31, 281–331.

Carrasco, M., and X. Chen (2002): "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18, 17–39.

Chalak, K., and H. White (2008): "Independence and Conditional Independence in Causal Systems," Discussion paper, University of California, San Diego.

Chen, S. X., and T. Huang (2007): "Nonparametric Estimation of Copula Functions for Dependent Modeling," *Canadian Journal of Statistics*, 35, 265–282.

Chen, X., and Y. Fan (2006a): "Estimation and Model Selection of Semiparametric Copula-based Multivariate Dynamic Models under Copula Misspecification," *Journal of Econometrics*, 135, 125–154.

——— (2006b): "Estimation of Copula-based Semiparametric Time Series Models," *Journal of Econometrics*, 130, 307–335.

Christie, A. (1982): "The Stochastic Behavior of Common Stock Variances- Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, 3, 145–166.

Comte, F., and E. Renault (1998): "Long Memory in Continuous Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291–323.

Deheuvels, P. (1979): *La Fonction de Dépendance Empirique et Ses Propriétés: Un Test Non-paramétrique D'indépendance* pp. 274–292. Bullettin de l'académie Royal de Belgique, Classe des Sciences.

Delgado, M., and W. Manteiga (2001): "Significance Testing in Nonparametric Regression based on the Bootstrap," *Annals of Statistics*, 29, 1469–1507.

Dufour, J.-M., R. Garcia, and A. Taamouti (2008): "Measuring Causality Between Volatility and Returns with High-Frequency Data," Discussion paper, Université de Montréal and Universidad Carlos III de Madrid.

Dufour, J.-M., and E. Renault (1998): "Short-Run and Long-Run Causality in Time Series: Theory," *Econometrica*, 66, 1099–1125.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Epps, T. W., and M. Epps (1976): "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implication for The Mixture-of-Distributions Hypothesis," *Econometrica*, 44, 305–322.

Fan, Y., and Q. Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semi-parametric Functional Forms," *Econometrica*, 64, 865–890.

Fan, Y., and Q. Li (1999): "Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Tests," *Journal of Nonparametric Statistics*, 10, 245–271.

Fan, Y., and Q. Li (2001): "Consistent Model Specification Tests: Nonparametric versus Bierens ICM tests," *Econometric Theory*, 16, 1016–1041.

Fan, Y., and O. Linton (2003): "Some Higher-order Theory for a Consistent Non-parametric Model Specification Test," *Journal of Statistical Planning and Inference*, 109, 125–154.

Fermanian, J., and O. Scaillet (2003): "Nonparametric Estimation of Copulas for Time Series," *Journal of Risk*, 5, 25–54.

Florens, J., and M. Mouchart (1982): "A Note on Non-causality," *Econometrica*, 50, 583–591.

Florens, J.-P., and D. Fougère (1996): "Non-causality in Continuous Time," *Econometrica*, 64, 1195–1212.

French, M., W. S., and R. Stambaugh (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–30.

Gallant, R., P. Rossi, and G. Tauchen (1992): "Stock Prices and Volume," *Review of Financial Studies*, 5, 199–242.

Gao, J., and I. Gijbels (2008): "Bandwidth Selection in Nonparametric Kernel Testing," *Journal of the American Statistical Association*, 103, 1584–1594.

Gervais, S., R. Kaniel, and D. Mingelgrin (2001): "The High-Volume Return Premium," *The Journal of Finance*, 56(3), 877–919.

Gijbels, I., and J. Mielniczuk (1990): "Estimating The Density of a Copula Function," *Communications in Statistics - Theory and Methods*, 19, 445–464.

Glosten, L. R., R. J., and D. E. Runkle (1993): "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

Gouriéroux, C., and C. Tenreiro (2001): "Local Power Properties of Kernel Based Goodness of Fit Tests," *Journal of Multivariate Analysis*, 78, 161–190.

Granger, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–459.

Hall, P. (1984): "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators," *Journal of Multivariate Analysis*, 14, 1–16.

Hiemstra, C., and J. Jones (1994): "Testing for Linear and Nonlinear Granger Causality In The Stock Price-Volume Relation," *Journal of Finance*, v49(5), 1639–1664.

Hoeffding, W. (1948): "A Non-parametric Test of Independence," *The Annals of Mathematical Statistics*, 19, 546–557.

Hong, Y., and H. White (2005): "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence," *Econometrica*, 73, 837–901.

Horowitz, J., and V. Spokoiny (2001): "Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternatives," *Econometrica*, 69, 599–631.

Hsiao, C. (1982): "Autoregressive Modeling and Causal Ordering of Economic Variables," *Journal of Economic Dynamics and Control*, 4, 243–259.

Huang, X., and G. Tauchen (2005): "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, 3(4), 456–499.

Karpoff, J. M. (1987): "The Relation Between Price Changes and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis*, v22(1), 109–126.

Kitamura, Y. (2001): "Asymptotic Optimality of Empirical Likelihood for Testing Moment restrictions," *Econometrica*, 69, 1661–1672.

Lee, S., and Y.-J. Whang (2009): "Nonparametric Tests of Conditional Treatment Effects," Discussion paper, University College London and Seoul National University.

Li, Q., E. Maasoumi, and J. Racine (2009): "A Nonparametric Test for Equality of Distributions with Mixed Categorical and Continuous Data," *Journal of Econometrics*, 148, 186–200.

Linton, O., and P. Gozalo (1997): "Conditional Independence Restrictions: Testing and Estimation," Discussion paper, Cowles Foundation Discussion Paper 1140.

Lütkepohl, H. (1993): *Testing for Causation between two Variables in Higher Dimensional VAR models in H. Schneeweiss and K. Zimmermann, eds, 'Studies in Applied Econometrics'*. Springer-Verlag, Heidelberg.

Meitz, M., and P. Saikkonen (2002): "Ergodicity, Mixing, and Existence of Moments of a Class of Markov Models with Applications to GARCH and ACD Models," *Econometric Theory*, 24, 1291–1320.

Morgan, I. (1976): "Stock Prices and Heteroskedasticity," *Journal of Business*, 49, 496–508.

Nelson, D. B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Paparoditis, E., and D. Politis (2000): "The Local Bootstrap for Kernel Estimators under General Dependence Conditions," *Annals of the Institute of Statistical Mathematics*, 52, 139–159.

Pindyck, R. (1984): "Risk, Inflation, and the Stock Market," *American Economic Review*, 74, 334–351.

Robinson, P. (1991): "Consistent Nonparametric Entropy-based Testing," *Review of Economic Studies*, 58, 437–453.

Rogalski, R. (1978): "The Dependence of Prices and Volumes," *The Review of Economics and Statistics*, 36, 268–274.

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.

Sancetta, A., and S. Satchell (2004): "The Bernstein Copula and its Applications to Modeling and Approximating of Multivariate Distributions," *Econometric Theory*, 20, 535–562.

Schwert, G. (1989): "Why Does Stock Market Volatility Change Over Time?," *Journal of Finance*, 44, 1115–1153.

Sims, C. (1972): "Money, Income and Causality," *American Economic Review*, 62, 540–552.

——— (1980a): "Comparison of Interwar and Postwar Business Cycles: Monetarism reconsidered," *American Economic Review*, 70, 250–257.

——— (1980b): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Skaug, H. J., and D. Tjostheim (1993): "A nonparametric test of serial independence based on the empirical distribution function," *Biometrika*, 80, 591–602.

Sklar, A. (1959): "Fonction de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

Su, L., and H. White (2003): "Testing Conditional Independence via Empirical Likelihood," Discussion Paper, University of California San Diego.

——— (2007): "A Consistent Characteristic Function-based test for Conditional Independence," *Journal of Econometrics*, 141, 807–834.

——— (2008): "A nonparametric Hellinger Metric test for Conditional Independence," *Econometric Theory*, 24, 1–36.

Tenreiro, C. (1997): "Loi Asymptotique des Erreurs Quadratiques Intégrées des Estimateurs a Noyau de la Densité et de la Régression sous des Conditions de Dépendance," *Portugaliae Mathematica*, 54, 197–213.

Tripathi, G., and Y. Kitamura (2003): "Testing Conditional Moment Restrictions," *Annals of Statistics*, 31, 2059–2095.

Turner, C., R. Startz, and C. Nelson (1989): "A Markov Model of Heteroskedasticity, Risk and Learning in the Stock Market," *Journal of Financial Economics*, 25, 3–22.

Westerfield, R. (1977): "The Distribution of Common Stock Price Changes: An Application of Transactions Time and Subordinated Stochastic Models," *Journal of Financial and Quantitative Analysis*, 12, 743–765.

White, H., and X. Lu (2008): "Granger Causality and Dynamic Structural Systems," Discussion paper, University of California, San Diego.

WIENER, N. (1956): *The Theory of Prediction* chap. 8. McGraw-Hill, New York.

YOSHIHARA, K. (1976): "Limiting Behaviour of U-statistics for Stationary Absolutely Regular Process," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 35, 237–252.