Journal of Multivariate Analysis 107 (2012) 210–226

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A conditional independence test for dependent data based on maximal conditional correlation

Yu-Hsiang Cheng, Tzee-Ming Huang *

Department of Statistics, National Chengchi University, 116 Taipei, Taiwan, ROC

ARTICLE INFO

Article history:
Received 29 April 2010
Available online 31 January 2012

AMS 2010 subject classifications:
62G10
62H20

Keywords:
Conditional independence test
α-mixing
Maximal conditional nonlinear correlation

ABSTRACT

In Huang (2010) [8], a test of conditional independence based on maximal nonlinear conditional correlation is proposed and the asymptotic distribution for the test statistic under conditional independence is established for IID data. In this paper, we derive the asymptotic distribution for the test statistic under conditional independence for α-mixing data. The results of simulation show that the test performs reasonably well for dependent data. We also apply the test to stock index data to test Granger noncausality between returns and trading volume.

1. Introduction

The testing of conditional independence is important in statistics; one interesting application of such testing is variable selection. For instance, consider the following regression problem:

$$Y = f(Z, X) + \epsilon, \tag{1}$$

where ϵ is independent of (Z, X) and f is a real-valued function. If Y and X are conditionally independent given Z, the variable X can be excluded from the model in (1).

Suppose that X, Y and Z are continuous random vectors of dimensions d_1, d_2 and d respectively. For testing whether X and Y are conditionally independent given Z, most tests in the literature deal with the case where the observations for (X, Y, Z) are IID. See, for example, [11,3,9,8], etc.

When the observations for (X, Y, Z) are weakly dependent, fewer tests are available in the literature. Su and White [14,15] developed nonparametric tests based on a weighted Hellinger distance between conditional densities or the difference between conditional characteristic functions. Bouezmarni et al. [1] also proposed a nonparametric test based on the Hellinger distance of copula densities.

In [14,15,1], one motivation for constructing conditional independence tests for dependent data is to test Granger noncausality, which, according to Florens and Mouchart [5] and Florens and Fougere [4], is a form of conditional independence. Specifically, a series $\{U_t\}$ does not Granger cause series $\{V_t\}$ if

$$V_t \perp (U_{t-1}, U_{t-2}, \ldots, U_{t-p}) | (V_{t-1}, V_{t-2}, \ldots, V_{t-p}) \quad \text{for every } p \geq 1,$$

where $\perp$ denotes an independent relationship.

* Corresponding author.
E-mail addresses: 96354501@nccu.edu.tw (Y.-H. Cheng), tmhuang@nccu.edu.tw (T.-M. Huang).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.01.017

In this paper, we consider Huang's test statistic and derive its asymptotic distribution for α-mixing data. In order to measure the conditional association between X and Y given Z, Huang [8] uses a measure called the maximal nonlinear conditional correlation, which is defined as

$$\sup_{f,g\in S_0^*} \mathrm{Corr}(f(X,Z), g(Y,Z)|Z), \tag{2}$$

where S_0^* is the collection of (f,g)'s such that $E(f^2(X,Z)) < \infty$ and $E(g^2(Y,Z)) < \infty$. Huang's test statistic is an estimator for a weighted average of estimators of maximal nonlinear conditional correlation at different evaluation points for the given variable Z. The test statistic also involves certain basis functions used to approximate the f and g in (2). We show that the asymptotic distribution of Huang's test statistic for α-mixing data is the same as that for IID data if the number of evaluation points and the number of basis functions are held constant.

This paper is organized as follows. In Section 2, we review the definition of maximal nonlinear conditional correlation and certain approximation results given in [8], and state the asymptotic properties of the test statistic that we derive under α-mixing condition. Some simulation results and an application are in Section 3. Proofs are given in Section 4.

2. Review and main results

In this section, we review the definition of the maximal nonlinear conditional correlation $\rho_1(X,Y|Z)$, the approximation of $\rho_1(X,Y|Z)$ and the proposed estimator for $\rho_1(X,Y|Z=z)$ in [8]. Then, we consider Huang's test statistic for testing $H_0 : \rho_1(X,Y|Z) = 0$ and present its asymptotic properties that we derive under α-mixing condition.

2.1. Definition, approximation, and estimation for maximal nonlinear conditional correlation

The maximal nonlinear conditional correlation $\rho_1(X,Y|Z)$ is essentially the maximum of $E(f(X,Z)g(Y,Z)|Z)$ over S_0, where S_0 is the collection of (f,g)'s that satisfy the following conditions:

$$\begin{aligned} &E(f^2(X,Z)|Z)I_{(0,\infty)}(E(f^2(X,Z)|Z)) = I_{(0,\infty)}(E(f^2(X,Z)|Z)) \\ &E(g^2(Y,Z)|Z)I_{(0,\infty)}(E(g^2(Y,Z)|Z)) = I_{(0,\infty)}(E(g^2(Y,Z)|Z)) \end{aligned} \tag{3}$$

and

$$E(f(X,Z)|Z) = E(g(Y,Z)|Z) = 0. \tag{4}$$

To avoid dealing with the existence of the maximum and the measurability of $\rho_1(X,Y|Z)$, in [8], $\rho_1(X,Y|Z)$ is defined as

$$\sup_{(f,g)\in S_0} E(f(X,Z)g(Y,Z)|Z),$$

where the supremum is defined as

$$\lim_{n\to\infty} E(\alpha_n(X,Z)\beta_n(Y,Z)|Z),$$

where $\{(\alpha_n, \beta_n)\}$ is a sequence in S_0 that satisfies the following conditions.

(i) The sequence $\{E(\alpha_n(X,Z)\beta_n(Y,Z)|Z)\}$ is non-decreasing.
(ii) For every $(f,g) \in S_0$,

$$E(f(X,Z)g(Y,Z)|Z) \le \lim_{n\to\infty} E(\alpha_n(X,Z)\beta_n(Y,Z)|Z).$$

To approximate

$$\rho_1(X,Y|Z) = \sup_{(f,g)\in S_0} E(f(X,Z)g(Y,Z)|Z),$$

we consider $S_{0,p,q}$: the collection of all (f,g)'s in S_0 such that f and g are in the spans of $\{\phi_{p,j} : 1 \le j \le p\}$ and $\{\psi_{q,k} : 1 \le k \le q\}$ respectively, when Z is given. That is,

$$f(X,Z) = \sum_{j=1}^{p} a_{p,j}(Z)\phi_{p,j}(X) \quad \text{for some } a_{p,j}(Z)\text{'s}$$

and

$$g(Y,Z) = \sum_{k=1}^{q} b_{q,k}(Z)\psi_{q,k}(Y) \quad \text{for some } b_{q,k}(Z)\text{'s}.$$

Suppose that the basis functions $\phi_{p,i}$'s and $\psi_{q,j}$'s are selected so that there exists basis functions $\theta_{r,k}$'s such that

$$\lim_{p,r\to\infty}\inf_{a(i,k)} E\left(\alpha(X,Z)-\sum_{1\le i\le p,1\le k\le r} a(i,k)\phi_{p,i}(X)\theta_{r,k}(Z)\right)^2=0 \quad (5)$$

and

$$\lim_{q,r\to\infty}\inf_{b(j,k)} E\left(\beta(Y,Z)-\sum_{1\le j\le q,1\le k\le r} b(j,k)\psi_{q,j}(Y)\theta_{r,k}(Z)\right)^2=0 \quad (6)$$

for every α and β such that $E(\alpha^2(X,Z))$ and $E(\beta^2(Y,Z))$ are finite. Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be the ranges of X, Y and Z respectively. Suppose that for each (p,q), there exist coefficients $a_{p,0,i}$'s and $b_{q,0,j}$'s such that

$$\sum_{1\le i\le p} a_{p,0,i}\phi_{p,i}(x)=1=\sum_{1\le j\le q} b_{q,0,j}\psi_{q,j}(y) \quad (7)$$

for every x in $\mathcal{X}$ and every y in $\mathcal{Y}$. Let

$$\rho_{p,q}(Z)=\max_{(f,g)\in S_{0,p,q}} E(f(X,Z)g(Y,Z)|Z).$$

Then, by Fact 2 in [8], $\rho_1(X,Y|Z)$ can be reasonably approximated by $\rho_{p,q}(Z)$ if p and q are large. The statement of the fact is given below.

Fact 1 (*Fact 2 in [8]*). *Suppose that (5)–(7) hold and $\{p_n\}$ and $\{q_n\}$ are sequences of positive integers that tend to ∞ as $n\to\infty$. Then*

$$\lim_{n\to\infty} E(|\rho_1(X,Y|Z)-\rho_{p_n,q_n}(Z)|)=0.$$

A remark follows.

- It is not difficult to find basis functions that satisfy (5)–(7). If $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are bounded regions in R^{d_1}, R^{d_2} and R^d respectively and the Lebesgue densities for (X,Z) and (Y,Z) are bounded, then $\phi_{p,i}$'s and $\psi_{q,j}$'s can be taken as B-spline basis functions on multidimensional intervals containing $\mathcal{X}$ and $\mathcal{Y}$ respectively, where the $\theta_{r,k}$'s can be taken as B-spline basis functions on a multidimensional interval containing $\mathcal{Z}$.

$\rho_{p,q}(Z)$ can be found as follows. First, we look for vectors $a_1=(a_{1,1}(Z),\ldots,a_{1,p}(Z))^T$ and $b_1=(b_{1,1}(Z),\ldots,b_{1,q}(Z))^T$ such that (a_1,b_1) is the pair (a,b) that maximizes $a^T\Sigma_{\phi,\psi,p,q}(Z)b$ subject to

$$a^T\Sigma_{\phi,p}(Z)a=1=b^T\Sigma_{\psi,p}(Z)b,$$

where

$$\Sigma_{\phi,p}(Z)=(E(\phi_{p,i}(X)\phi_{p,j}(X)|Z)-E(\phi_{p,i}(X)|Z)E(\phi_{p,j}(X)|Z))_{p\times p},$$
$$\Sigma_{\psi,q}(Z)=(E(\psi_{q,i}(Y)\psi_{q,j}(Y)|Z)-E(\psi_{q,i}(Y)|Z)E(\psi_{q,j}(Y)|Z))_{q\times q},$$

and

$$\Sigma_{\phi,\psi,p,q}(Z)=(E(\phi_{p,i}(X)\psi_{q,j}(Y)|Z)-E(\phi_{p,i}(X)|Z)E(\psi_{q,j}(Y)|Z))_{p\times q}.$$

Take

$$f_1(X,Z)=\sum_{j=1}^{p} a_{1,j}(Z)(\phi_{p,j}(X)-E(\phi_{p,j}(X)|Z))$$

and

$$g_1(Y,Z)=\sum_{k=1}^{q} b_{1,k}(Z)(\psi_{q,j}(X)-E(\psi_{q,j}(Y)|Z)).$$

Then, $E(f_1(X,Z)g_1(Y,Z)|Z)=\rho_{p,q}(Z)$.

For $z\in\mathcal{Z}$, let $\hat{\Sigma}_{\phi,\psi,p,q}(z)$, $\hat{\Sigma}_{\phi,p}(z)$ and $\hat{\Sigma}_{\psi,q}(z)$ be the kernel estimators of $\Sigma_{\phi,\psi,p,q}(z)$, $\Sigma_{\phi,p}(z)$ and $\Sigma_{\psi,q}(z)$ respectively; in other words, every element $E(g(X,Y)|Z=z)$ in $\Sigma_{\phi,\psi,p,q}(z)$, $\Sigma_{\phi,p}(z)$ and $\Sigma_{\psi,q}(z)$ is estimated by

$$\hat{E}(g(X,Y)|Z=z)=\frac{\sum_{t=1}^{n} g(X_t,Y_t)k_0((Z_t-z)/h)}{\sum_{t=1}^{n} k_0((Z_t-z)/h)} \quad (8)$$

in $\hat{\Sigma}_{\phi,\psi,p,q}(z)$, $\hat{\Sigma}_{\phi,p}(z)$ and $\hat{\Sigma}_{\psi,q}(z)$, where k_0 is a kernel function defined on R^d and $h > 0$. Then, we use $\hat{\rho}_{p,q}(z) = \max_{a,b} a^T \hat{\Sigma}_{\phi,\psi,p,q}(z)b$ for estimating $\rho_{p,q}(z)$, where all pairs (a, b) satisfy

$$a^T \hat{\Sigma}_{\phi,p}(z)a = 1 = b^T \hat{\Sigma}_{\psi,q}(z)b.$$

Henceforth, the estimator $\hat{\rho}_{p,q}(z)$ will be abbreviated as $\hat{\rho}(z)$ for each z in $\mathcal{Z}$.

2.2. A test for conditional independence and relative asymptotic properties

The conditional independence test that we use in this paper is based on $\hat{\rho}^2(z)$ at different z's. Since each $\hat{\rho}(z)$ is determined by the kernel estimators of certain conditional expectations, we first derive their joint asymptotic distribution. Then, we use $\sum_{i=1}^{k} \hat{f}_Z(z_i)\hat{\rho}^2(z_i)$'s as our test statistic and establish its consistency and asymptotic distribution. Here the z_i's are selected points in $\mathcal{Z}$ and

$$\hat{f}_Z(\cdot) = \frac{\sum_{t=1}^{n} k_0((Z_t - \cdot)/h)}{nh^d}$$

is the kernel density estimator of f_Z: the Lebesgue pdf of Z. In order to avoid dealing with the boundary bias problem in kernel estimation, we consider a set $\mathcal{Z}^0$ that is contained in the interior of $\mathcal{Z}$ so that points in $\mathcal{Z}^0$ are away from the boundary of $\mathcal{Z}$, and choose the z_i's from $\mathcal{Z}^0$.

Our first result is with regard to the joint asymptotic distribution of kernel estimators of some conditional expectations. In order to describe the assumptions, we first review the definition for α-mixing coefficients. For a strictly stationary process $\{U_t\}$, let $\mathcal{F}_a^b$ denote the σ-algebra generated by $(U_a, \ldots, U_b)$. Then, the α-mixing coefficient at lag s for $\{U_t\}$ is

$$\sup\left\{|P(A \cap B) - P(A)P(B)| : -\infty < t < \infty, A \in \mathcal{F}_{-\infty}^{t}, B \in \mathcal{F}_{t+s}^{\infty}\right\}.$$

$\{U_t\}$ is considered to be α-mixing if its α-mixing coefficient at lag s tends to 0 as s tends to ∞. Let $\alpha(s)$ denote the α-mixing coefficient at lag s for the process $\{(X_t, Y_t, Z_t)\}$. Our assumptions are provided below.

(S0) The basis functions $\phi_{p,1}, \ldots, \phi_{p,p}$ and $\psi_{q,1}, \ldots, \psi_{q,q}$ are bounded and (5)–(7) hold. For the sake of brevity, $\phi_{p,1}, \ldots, \phi_{p,p}$ and $\psi_{q,1}, \ldots, \psi_{q,q}$ will be abbreviated as $\phi_1, \ldots, \phi_p$ and $\psi_1, \ldots, \psi_q$ respectively hereafter.

(S1) $\{(X_t, Y_t, Z_t) \in R^{d_1+d_2+d}, t \geq 0\}$ is a strictly stationary α-mixing process that satisfies $\alpha(\tau) = O(\tau^{-(1+\epsilon)})$, where $\epsilon > \max(1, d/2)$, d_1, d_2 and d denote the dimensions of X_t, Y_t and Z_t respectively.

(S2) Suppose that there exist $\mathcal{Z}^0$: an open subset of the interior of $\mathcal{Z}$ and μ: σ-finite measure such that for every $z \in \mathcal{Z}^0$, the conditional distribution of (X, Y) given $Z = z$ has a pdf $f(\cdot|z)$ with respect to μ. Further, $f(x, y|z)$ and $f_Z(z)$ are twice differentiable with respect to z on $\mathcal{Z}^0$.

(S3) There exists a function h on $\mathcal{X} \times \mathcal{Y}$ such that

$$\sup_{z\in\mathcal{Z}^0} \max\left(|f(x, y|z)|, \max_{1\leq i\leq d}\left|\frac{\partial}{\partial z_i} f(x, y|z)\right|, \max_{1\leq i,j\leq d}\left|\frac{\partial^2}{\partial z_i \partial z_j} f(x, y|z)\right|\right) \leq h(x, y)$$

and $\int h(x, y)d\mu(x, y) < \infty$.

(S4) There exist constants c_0 and c_1 such that

$$\sup_{z\in\mathcal{Z}^0} \max\left(|f_Z(z)|, \max_{1\leq i\leq d}\left|\frac{\partial}{\partial z_i} f_Z(z)\right|, \max_{1\leq i,j\leq d}\left|\frac{\partial^2}{\partial z_i \partial z_j} f_Z(z)\right|\right) \leq c_0$$

and $1/f_Z(z) \leq c_1$ for $z \in \mathcal{Z}^0$.

(S5) k^* is a kernel function defined on R^1, and k_0 is a product kernel on R^d that satisfies

$$k_0(v_1, v_2, \ldots, v_d) = k^*(v_1)k^*(v_2)\cdots k^*(v_d),$$

$k^* \geq 0$, $\sup_v k^*(v) < \infty$, $\int k^*(v)dv = 1$, $\int vk^*(v)dv = 0$, $\int v(k^*(v))^2dv = 0$ and $\kappa_2 = \int v^2k^*(v)dv < \infty$.

(S6) As $n \to \infty$, the bandwidth $h \to 0$, $nh^d \to \infty$ and $nh^{d+4} \to 0$.

Under the above conditions, the joint asymptotic distribution of kernel estimators of conditional expectations can be established, as stated in Lemma 1. The proof for Lemma 1 is provided in Section 4.1.

Lemma 1. *Suppose that conditions* (S1)–(S6) *hold. Suppose that* $g_1, g_2, \ldots, g_m$ *are bounded functions defined on* $\mathcal{X} \times \mathcal{Y}$*. Suppose* $z_1, \ldots, z_k$ *are distinct points in* $\mathcal{Z}^0$*. For* $i = 1, \ldots, k$*, let*

$$\hat{g}_j(z_i) = \frac{\sum_{t=1}^{n} g_j(X_t, Y_t)k_0((Z_t - z_i)/h)}{\sum_{t=1}^{n} k_0((Z_t - z_i)/h)}$$

be the kernel estimator of $g_j^*(z_i) \equiv E(g_j(X, Y)|Z = z_i)$. *Further, let*

$$B_{s,j}(z_i) = \frac{\kappa_2}{2}(f_Z(z_i)g_{j,ss}^*(z_i) + 2f_s(z_i)g_{j,s}^*(z_i)) \tag{9}$$

and

$$\mathbf{W}_{j,n}(z_i) = \sqrt{nh^d}\left(\hat{g}_j(z_i) - g_j^*(z_i) - h^2\sum_{s=1}^{d} B_{s,j}(z_i)/f_Z(z_i)\right)$$

for $1 \le i \le k$ *and* $1 \le j \le m$, *where* $g_{j,s}^*$ *and* $g_{j,ss}^*$ *denote the first and the second partial derivatives of* g_j^* *with respect to the s-th component respectively and* f_s *denotes the first partial derivative of* f_Z *with respect to the s-th component. Let*

$$u_{j,t} = g_j(X_t, Y_t) - g_j^*(Z_t),$$
$$c_{jj^*}(z_i) = E(u_{j,1}u_{j^*,1}|Z_1 = z_i),$$
$$\sigma_j^2(z_i) = E(u_{j,1}^2|Z_1 = z_i),$$

and

$$\mathbf{W}_n = (\mathbf{W}_{1,n}(z_1), \ldots, \mathbf{W}_{1,n}(z_k), \ldots, \mathbf{W}_{m,n}(z_1), \ldots, \mathbf{W}_{m,n}(z_k))^T.$$

Then, $\mathbf{W}_n$ *converges in distribution to a random vector*

$$(Z_{1,1}^*, \ldots, Z_{k,1}^*, \ldots, Z_{1,m}^*, \ldots, Z_{k,m}^*)^T \equiv Z^*,$$

where Z^* *is multivariate normal with mean* 0 *and for* $1 \le i, i^* \le k$ *and* $1 \le j, j^* \le m$,

$$\mathrm{Cov}(Z_{i,j}^*, Z_{i^*,j^*}^*) = \begin{cases} \kappa^d\sigma_j^2(z_i)/f_Z(z_i) & \text{if } i = i^* \text{ and } j = j^*; \\ \kappa^d c_{jj^*}(z_i)/f_Z(z_i) & \text{if } i = i^* \text{ and } j \ne j^*; \\ 0 & \text{if } i \ne i^*, \end{cases}$$

where $\kappa = \int (k^*(v))^2 dv$.

Now, suppose that the basis functions ϕ_l's and ψ_{m^*}'s are linearly independent. For the sake of convenience, for $z \in \{z_1, \ldots, z_k\}$, we apply certain linear transformations to ϕ_l's and ψ_{m^*}'s to obtain new basis functions ϕ_l^*'s and $\psi_{m^*}^*$'s (the $\hat{\rho}(z)$ remains unchanged under such transformations). Take $g_1(X, Y), \ldots, g_m(X, Y)$ to be the functions $\phi_l^*(X)\phi_{l'}^*(X)$, $\phi_l^*(X)\psi_{m^*}^*(Y)$ and $\psi_{m^*}^*(Y)\psi_{m'}^*(Y)$, where $1 \le l \le l' \le p$ and $1 \le m^* \le m' \le q$. Then, the consistency of $\hat{\rho}(z)$ can be established and we have Theorem 1. The proof for Theorem 1 is provided in Section 4.2.

Theorem 1. *Suppose that conditions* (S0)–(S6) *hold and the basis functions* ϕ_l*'s and* ψ_{m^*}*'s are linearly independent. Suppose* $z_1, \ldots, z_k$ *are distinct points in* $\mathcal{Z}^0$. *Then,*

$$\sum_{i=1}^{k}\left(\hat{\rho}^2(z_i) - \rho_{p,q}^2(z_i)\right)^2 = O_p\left(\frac{1}{nh^d} + h^4\right)$$

and

$$\left(\sum_{i=1}^{k}\hat{f}_Z(z_i)\hat{\rho}^2(z_i) - \sum_{i=1}^{k} f_Z(z_i)\rho_{p,q}^2(z_i)\right)^2 = O_p\left(\frac{1}{nh^d} + h^4\right).$$

The following theorem states the approximate distribution of the statistic $\sum_{i=1}^k \hat{f}_Z(z_i)\hat{\rho}^2(z_i)$ when X and Y are conditionally independent given Z.

Theorem 2. *Suppose that the conditions in Theorem 1 hold and X and Y are conditionally independent given Z. Then,*

$$\frac{nh^d}{\kappa^d}\sum_{i=1}^{k}\hat{f}_Z(z_i)\hat{\rho}^2(z_i) \text{ converges in distribution to } \sum_{i=1}^{k}\lambda_i$$

as n tends to ∞, *where the* λ_i*'s are IID and have the same distribution as the largest eigenvalue of a matrix* CC^T, *where C is a* $(p-1) \times (q-1)$ *matrix whose elements are IID* $N(0, 1)$.

The proof of Theorem 2 is provided in Section 4.3. Theorem 2 is similar to Theorem 3.2 given in [8]. The main difference between the two is that Theorem 2 can be applied to α-mixing data. In addition, p and q are held fixed in Theorem 2, while they are allowed to depend on n and tend to ∞ as n tends to ∞ in Theorem 3.2 in [8].

According to Theorem 2, a test that rejects H_0 if

$$\frac{nh^d}{\kappa^d}\sum_{i=1}^{k}\hat{f}(z_i)\hat{\rho}^2(z_i) \geq F^*_{1-\alpha} \tag{10}$$

is of approximate level α, where F^* is the distribution function of $\sum_{i=1}^k \lambda_i$ and $F^*_{1-\alpha}$ is the $1-\alpha$ quantile of F^*. Theorem 3 states that the test with rejection region in (10) is consistent if $\rho_{p,q}(z_i) > 0$ for some selected z_i. The proof for this theorem is provided in Section 4.4.

Theorem 3. *Suppose that the conditions in Theorem 1 hold and $\rho_{p,q}(z_i) > 0$ for some $i \in \{1, \ldots, k\}$. Then, for $0 < \alpha < 1$, the probability that (10) holds tends to 1 as $n \to \infty$.*

3. Simulation studies and application to S&P500 index data

3.1. Simulation studies

In this section, we conduct several simulation studies for illustrating the performance of our test. The data generating processes, labeled Data1–Data13, are described below. In order to make our simulation results comparable with those of the test proposed by Su and White [15], some of our data generating processes (Data1–Data10) are the same as theirs. Throughout the description for Data1–Data10, $(\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t})$ are IID $N(0, I_3)$.

Data1: $(X_t, Y_t, Z_t) = (\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t})$.
Data2: $X_t = 0.5X_{t-1} + \epsilon_{1,t}$, $Y_t = 0.5Y_{t-1} + \epsilon_{2,t}$ and $Z_t = X_{t-1}$.
Date3: $X_t = \epsilon_{1,t}\sqrt{0.01 + 0.5X^2_{t-1}}$, $Y_t = 0.5Y_{t-1} + \epsilon_{2,t}$ and $Z_t = X_{t-1}$.
Data4: $X_t = \epsilon_{1,t}\sqrt{h_{1,t}}$, $Y_t = \epsilon_{2,t}\sqrt{h_{2,t}}$, $Z_t = X_{t-1}$, $h_{1,t} = 0.01 + 0.9h_{1,t-1} + 0.05X^2_{t-1}$ and $h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Y^2_{t-1}$.
Data5: $X_t = 0.5X_{t-1} + 0.5Y_t + \epsilon_{1,t}$, $Y_t = 0.5Y_{t-1} + \epsilon_{2,t}$ and $Z_t = X_{t-1}$.
Data6: $X_t = 0.5X_{t-1} + 0.5Y^2_t + \epsilon_{1,t}$, $Y_t = 0.5Y_{t-1} + \epsilon_{2,t}$ and $Z_t = X_{t-1}$.
Data7: $X_t = 0.5X_{t-1}Y_t + \epsilon_{1,t}$, $Y_t = 0.5Y_{t-1} + \epsilon_{2,t}$ and $Z_t = X_{t-1}$.
Data8: $X_t = 0.5X_{t-1} + 0.5Y_t\epsilon_{1,t}$, $Y_t = 0.5Y_{t-1} + \epsilon_{2,t}$ and $Z_t = X_{t-1}$.
Data9: $X_t = \epsilon_{1,t}\sqrt{0.01 + 0.5X^2_{t-1} + 0.25Y^2_t}$, $Y_t = 0.5Y_{t-1} + \epsilon_{2,t}$ and $Z_t = X_{t-1}$.
Data10: $X_t = \epsilon_{1,t}\sqrt{h_{1,t}}$, $Y_t = \epsilon_{2,t}\sqrt{h_{2,t}}$, $Z_t = X_{t-1}$, $h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4X^2_{t-1} + 0.5Y^2_t$ and $h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.5Y^2_t$.
Data11: $(X_t, Y_t, Z_t) = (\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t})$, where $(\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t})$ are IID $LN(0, I_3)$.
Data12: $X_t = \epsilon_{1,t}\epsilon_{1,t-1}$, $Y_t = \epsilon_{2,t}\epsilon_{2,t-1}$ and $Z_t = X_{t-1}$, where $(\epsilon_{1,t}, \epsilon_{2,t})$ are IID $LN(0, I_2)$.
Data13: $X_t = \epsilon_{1,t}\epsilon_{2,t-1}$, $Y_t = \epsilon^2_{1,t}\epsilon_{2,t-1}$ and $Z_t = \epsilon_{2,t-1}$, where $(\epsilon_{1,t}, \epsilon_{2,t})$ are IID $LN(0, I_2)$.

Here, Data1–Data4, Data11 and Data12 are used for examining the level of the test, and Data5–Data10 and Data13 are used for checking the power.

3.1.1. Simulation studies based on asymptotic distribution of the test statistic

We first apply our test using the asymptotic distribution of the test statistic.

Parameter set-up: in order to apply our test, certain parameters need to be specified, including the kernel function k^*, the kernel bandwidth h and the basis functions. For the sake of simplicity, in all the simulation experiments, we take the kernel bandwidth h to be $cn^{-0.25}$, where n is the sample size and $c \in \{0.5, 1, 1.5, 2\}$; we use the following kernel function:

$$k^*(x) = \begin{cases} 1 - x & \text{if } 0 \leq x \leq 1; \\ x + 1 & \text{if } -1 \leq x < 0. \end{cases}$$

In addition, the basis functions $\phi^*_1, \ldots, \phi^*_p$ and $\psi^*_1, \ldots, \psi^*_q$ are selected in the following manner. For $i = 1, \ldots, p$ and $j = 1, \ldots, q$, let

$$\phi_i(x) = \begin{cases} 1 & \text{if } \dfrac{i-1}{p} \leq x < \dfrac{i}{p}; \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

and

$$\psi_j(y) = \begin{cases} 1 & \text{if } \dfrac{j-1}{q} \leq y < \dfrac{j}{q}; \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

Table 1
Power results for different c's when $n = 500$ and $n = 1000$.

	$n = 500$				$n = 1000$			
	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 2$	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 2$
Data1	0.030	0.039	0.053	0.071	0.043	0.048	0.057	0.074
Data2	0.030	0.041	0.058	0.074	0.033	0.048	0.060	0.069
Data3	0.032	0.042	0.055	0.080	0.038	0.049	0.055	0.070
Data4	0.038	0.044	0.057	0.075	0.042	0.048	0.057	0.066
Data5	0.951	1	1	1	1	1	1	1
Data6	0.898	1	1	1	0.997	1	1	1
Data7	0.918	1	1	1	0.985	1	1	1
Data8	0.995	1	1	1	1	1	1	1
Data9	0.725	0.991	1	1	0.993	1	1	1
Data10	0.374	0.817	0.959	0.986	0.819	0.996	1	1
Data11	0.036	0.050	0.062	0.079	0.035	0.042	0.049	0.059
Data12	0.036	0.051	0.055	0.072	0.041	0.041	0.053	0.065
Data13	1	1	1	1	1	1	1	1

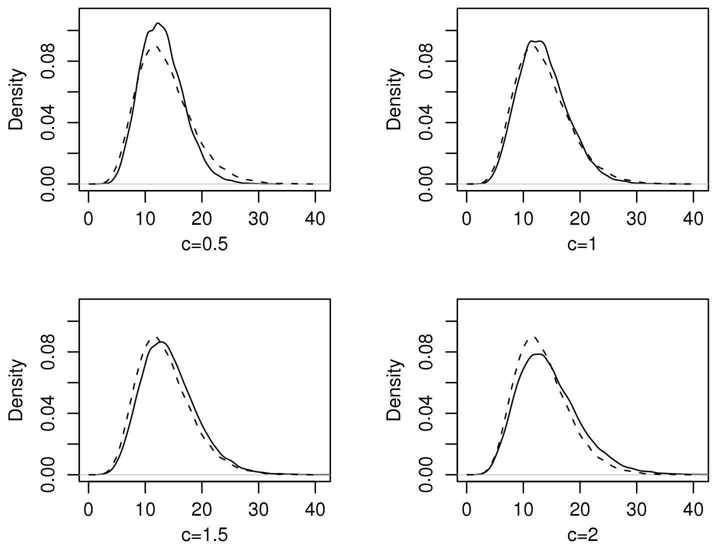

Fig. 1. Exact distribution (solid line) versus asymptotic distribution (dashed line) of the test statistic with different bandwidth choices ($h = cn^{-1/4}$).

where $p = q = 4$. Since the basis functions are defined on $[0, 1]$, we transform the data $(X_t, Y_t, Z_t)_{t=1}^n$ into $(F_1(X_t), F_2(Y_t), F_3(Z_t))_{t=1}^n$ before using the test, where F_1, F_2 and F_3 denote the empirical CDF's of $\{X_t\}_{t=1}^n$, $\{Y_t\}_{t=1}^n$ and $\{Z_t\}_{t=1}^n$ respectively. For the choice of the evaluation points, we take $z_1 = 0.78n^{-0.25} \equiv h_0$ and $z_i = z_{i-1} + 2h_0$ if $i \geq 2$ and $z_i \leq 1 - h_0$.

Table 1 shows that the levels of the test are less than 0.05 for $c = 0.5$ and $c = 1$ and the powers of the test are larger for larger c's. It seems that when $c = 1$, the levels of the test are close to 0.05 and the power performance is fine.

3.1.2. *Simulation studies based on local bootstrap*

The test based on asymptotic distribution of the test statistic does not work well for small sample sizes. Fig. 1 shows that the distribution of the test statistic and the asymptotic distribution are quite different for Data11 when $n = 100$. For Data1–Data4 and Data12, we find similar patterns. When $n = 200$, the difference between the distribution of the test statistic and the asymptotic distribution become smaller but is still visible.

To apply our test for small sample sizes, we consider the local bootstrap procedure proposed by Paparoditis and Politis [13]. The local bootstrap procedure is described below. For a given sample $\{(X_t, Y_t, Z_t)\}_{t=1}^n$, a local bootstrap sample $\{(X_t^*, Y_t^*, Z_t^*)\}_{t=1}^n$ is generated according to the following steps.

(a) Draw a random sample $(Z_1^*, Z_2^*, \ldots, Z_n^*)$ from the empirical cumulative distribution function $\hat{F}_Z$, where

$$\hat{F}_Z(z) = \frac{1}{n}\sum_{t=1}^{n} I_{(-\infty, Z_t]}(z).$$

Table 2
Power comparison between Tests 1 and 2 when $n = 100$.

	Data1	Data2	Data3	Data4	Data5	Data6	Data7
Test 2, $c^* = 1$	0.096	0.060	0.048	0.072	0.668	0.756	0.388
Test 2, $c^* = 2$	0.072	0.036	0.072	0.048	0.952	0.944	0.576
Test 1, $c = 0.5$	0.045	0.061	0.046	0.062	0.525	0.479	0.265
Test 1, $c = 1$	0.046	0.050	0.050	0.047	0.746	0.717	0.400
Test 1, $c = 1.5$	0.040	0.052	0.056	0.055	0.814	0.779	0.329
Test 1, $c = 2$	0.041	0.050	0.053	0.062	0.852	0.793	0.218
	Data8	Data9	Data10	Data11	Data12	Data13	
Test 2, $c^* = 1$	0.860	0.828	0.680	0.034	0.043	0.589	
Test 2, $c^* = 2$	0.940	0.988	0.912	0.022	0.022	0.859	
Test 1, $c = 0.5$	0.692	0.357	0.195	0.058	0.050	1	
Test 1, $c = 1$	0.873	0.566	0.320	0.049	0.048	1	
Test 1, $c = 1.5$	0.889	0.618	0.341	0.049	0.041	1	
Test 1, $c = 2$	0.860	0.631	0.348	0.046	0.045	1	

(b) For $1 \leq t \leq n$, we draw X_t^* and Y_t^* independently from the empirical cumulative distribution functions $\hat{F}_{X|Z=Z_t^*}$ and $\hat{F}_{Y|Z=Z_t^*}$ respectively, where

$$\hat{F}_{X|Z=Z_t^*}(x) = \frac{\sum_{t=1}^{n} k^*((Z_t^* - Z_t)/b) I_{(-\infty, X_t]}(x)}{\sum_{t=1}^{n} k^*((Z_t^* - Z_t)/b)}$$

and

$$\hat{F}_{Y|Z=Z_t^*}(y) = \frac{\sum_{t=1}^{n} k^*((Z_t^* - Z_t)/b) I_{(-\infty, Y_t]}(y)}{\sum_{t=1}^{n} k^*((Z_t^* - Z_t)/b)}.$$

Here, the bandwidth b is taken to be $n^{-0.2}$ and the kernel function k^* is the probability density function for $N(0, 1)$.

In order to determine the rejection region for a given sample, we repeat the above procedure to obtain bootstrap resamples and compute the test statistic $nh^d \kappa^{-d} \sum_{i=1}^{k} \hat{f}(z_i) \hat{\rho}^2(z_i)$ for the original sample and each local bootstrap resample. For a given level α, if the test statistic based on the given sample is larger than the $(1 - \alpha)$ quantile of the test statistics that are computed based on the local bootstrap resamples, we reject the conditional independence hypothesis at level α. The purpose of using the local bootstrap procedure is to generate a resample $\{(X_t^*, Y_t^*, Z_t^*)\}_{t=1}^n$ such that the distribution of Z^*, the conditional distributions of X^* given $Z^* = z$ and Y^* given $Z^* = z$ are close to the distribution of Z, the conditional distributions of X given $Z = z$ and Y given $Z = z$ respectively. In addition, since X_t^* and Y_t^* are generated independently given $Z_t^* = z$, they are conditionally independent given $Z_t^* = z$, irrespective of whether or not X and Y are conditionally independent given Z.

In these simulation studies, we choose the basis functions in (11) and (12) with $p = q = 5$. The evaluation points are $\{0.2, 0.4, 0.6, 0.8\}$, and the kernel bandwidth h to be $cn^{-0.25}$, where n is the sample size and $c \in \{0.5, 1, 1.5, 2\}$.

Finally, we present a few experimental results of our test (Test 1) and Su and White's test (Test 2). For Test 2, we run the simulations for Data11–Data13 with the bandwidth $h_n = c^* n^{-1/8.5}$, where $c^* = 1$ or 2. Each power estimate is based on 3000 repetitions, where 1000 local bootstrap resamples are used in each repetition. For the sake of comparison, we also list some power estimates for Test 2 for Data1–Data10, which are taken directly from [15]. They use 250 repetitions with 200 local bootstrap resamples for each repetition.

Tables 2 and 3 indicate the level and power estimates for Test 1 and Test 2 at significance level 5% when the sample sizes are 100 and 200 respectively.

3.2. Application to S&P500 index data

In this section, we apply the linear Granger causality test (hereafter denoted by Test LIN) and our conditional independence test (Test 1) in order to check the interaction between returns and volume for S&P500 index data at one day lag. There are 2514 observations for daily index returns and trading volume from January 2000 to December 2009, taken from Yahoo Finance. Here, the return for day t is defined as

$$R_t = 100 \log\left(\frac{P_t}{P_{t-1}}\right),$$

Table 3
Power comparison between Tests 1 and 2 when $n = 200$.

	Data1	Data2	Data3	Data4	Data5	Data6	Data7
Test 2, $c^* = 1$	0.064	0.052	0.080	0.080	0.900	0.960	0.596
Test 2, $c^* = 2$	0.044	0.060	0.056	0.048	1	1	0.864
Test 1, $c = 0.5$	0.040	0.061	0.036	0.055	0.827	0.830	0.488
Test 1, $c = 1$	0.049	0.051	0.057	0.054	0.982	0.983	0.831
Test 1, $c = 1.5$	0.046	0.048	0.049	0.053	0.995	0.989	0.827
Test 1, $c = 2$	0.045	0.045	0.047	0.057	0.997	0.995	0.735

	Data8	Data9	Data10	Data11	Data12	Data13
Test 2, $c^* = 1$	0.992	0.968	0.880	0.031	0.036	0.347
Test 2, $c^* = 2$	1	1	0.996	0.025	0.032	0.872
Test 1, $c = 0.5$	0.988	0.730	0.392	0.062	0.062	1
Test 1, $c = 1$	1	0.947	0.679	0.048	0.047	1
Test 1, $c = 1.5$	1	0.968	0.738	0.051	0.043	1
Test 1, $c = 2$	1	0.971	0.745	0.058	0.037	1

Table 4
p-values for Test LIN and Test 1 for testing the relationship between returns and volume changes.

H_0	$R_{t-1} \nRightarrow V_t^*$	$V_{t-1}^* \nRightarrow R_t$
Test LIN	0.000	0.804
Test 1	0.001	0.032

where P_t is the index value for day t. Moreover, the trading volume for day t (in dollars), denoted by V_t, is transformed into

$$V_t^* = \log\left(\frac{V_t}{V_{t-1}}\right).$$

The above transformations are commonly used in the analysis for financial data; for example, see [7,1]. The augmented Dickey–Fuller test reveals that the series $\{R_t\}$ and $\{V_t^*\}$ are stationary.

In order to examine whether $\{R_t\}$ is useful for predicting $\{V_t^*\}$, we consider the effects up to lag 1. Specifically, we test

$$H_0 : V_t^* \perp R_{t-1} | V_{t-1}^* \tag{13}$$

using Test 1. For Test LIN, it is assumed that

$$E(V_t^* | R_{t-1}, V_{t-1}^*) = a_1 R_{t-1} + b_1 V_{t-1}^*$$

and the null hypothesis is

$$H_0 : a_1 = 0. \tag{14}$$

We use the notation $R_{t-1} \nRightarrow V_t^*$ to denote the relation expressed in (13) or (14). The notation $V_{t-1}^* \nRightarrow R_t$ is defined analogously.

The p-values for Test LIN and Test 1 are provided in Table 4. For Test 1, we use the same parameter set-up as in Section 3.1.1 and find both the return-to-volume and volume-to-return relationships are significant at the 5% level. However, for Test LIN, the volume-to-return relationship is not significant. These findings are consistent with the results obtained in [7,1].

To illustrate the implementation of our test for the $d > 1$ case, we also apply the test to test

$$H_0 : V_t^* \perp (R_{t-1}, R_{t-2}) | (V_{t-1}^*, V_{t-2}^*) \tag{15}$$

and

$$H_0 : R_t \perp (V_{t-1}^*, V_{t-2}^*) | (R_{t-1}, R_{t-2}). \tag{16}$$

The empirical CDF transforms are applied component-wisely. For instance, we transform $(V_{t-1}^*, V_{t-2}^*,)_{t=4}^n$ into $(F_1(V_{t-1}^*), F_2(V_{t-2}^*))_{t=4}^n$, where $n = 2512$ and F_i is the empirical CDF of V_{t-i}^* for $i = 1, 2$. For the basis functions, we use 4 basis functions on $[0, 1]$: $\phi_1, \ldots, \phi_4$ and 4 basis functions $\psi_1, \ldots, \psi_4$ on $[0, 1]^2$, where $\phi_1, \ldots, \phi_4$ are given in (11) with $p = 4$, $\psi_1(y_1, y_2) = I_{[0,0.5)}(y_1) I_{[0,0.5)}(y_2)$, $\psi_2(y_1, y_2) = I_{[0,0.5)}(y_1) I_{[0.5,1)}(y_2)$, $\psi_3(y_1, y_2) = I_{[0.5,1)}(y_1) I_{[0,0.5)}(y_2)$ and $\psi_4 = 1 - \psi_1 - \psi_2 - \psi_3$. Here $I_A(\cdot)$ denotes the indicator function on A. In addition, the kernel bandwidth is $cn^{-1/(d+\delta)}$ with $c = 1.4$ and $\delta = 2.4$. The evaluation points are all the points in $S_{h_0}^2$, where $S_{h_0} = \{(2k-1)h_0 : k \text{ is an integer }\} \cap [h_0, 1 - h_0]$ and $h_0 = 0.78 n^{-1/(d+\delta)}$. Here c and δ are selected from certain candidate values so that the levels of the test are close to 0.05 when the data are IID $U(0, 1)$. The p-values for (15) and (16) are 0.017 and 0.370 respectively.

It has been brought to our attention by an anonymous referee that our test rejected

$$H_0 : R_t \perp V_{t-1}^* | R_{t-1},$$

but not the H_0 in (16). A possible explanation for this result is that the impact of V^*_{t-1} on R_t given R_{t-1} can be explained by R_{t-2}. To check on this conjecture, we apply our test to test

$$H_0 : R_t \perp V^*_{t-1} | (R_{t-1}, R_{t-2}) \tag{17}$$

and the test does not reject the H_0 in (17), which supports our conjecture.

Some remarks on the implementation of the test.

- It is recommended to choose evaluation points so that two evaluation points, z_i and z_j, are at least $2h$ away (for each component) when a compact kernel supported on $[-1, 1]^d$ is used. In such case, the $\hat{\rho}(z_i)$ and $\hat{\rho}(z_j)$ are independent for IID data, which makes the distribution of the test statistic close to the derived asymptotic distribution. Since $nh^d \to \infty$, h cannot be too small, which implies that the number of evaluation points cannot be too large.
- We apply empirical CDF transforms to our data so that the distribution of each component of X, Y and Z is supported on $[0, 1]$. The transforms are data dependent and it is not clear whether the transformed data can be treated as if they were transformed by the true underlying CDF. The simulation results are fine, but further investigation is needed.

4. Proofs

In this section, we give proofs for Theorems 1–3 and Lemma 1. Before giving the proofs, we first define and recall some notations. Recall that k^* is a kernel on R^1 and k_0 is a product kernel on R^d defined by

$$k_0(v_1, v_2, \dots, v_d) = k^*(v_1)k^*(v_2)\cdots k^*(v_d),$$

$$\kappa = \int (k^*(v))^2 dv$$

and $\kappa_2 = \int v^2 k^*(v)dv$.

For a $(p+q) \times (p+q)$ matrix V_0, let $g_{1,1}(V_0), g_{1,2}(V_0), g_{2,1}(V_0)$ and $g_{2,2}(V_0)$ denote the matrices of dimensions $p \times p$, $p \times q$, $q \times p$, $q \times q$ respectively such that

$$V_0 = \begin{pmatrix} g_{1,1}(V_0) & g_{1,2}(V_0) \\ g_{2,1}(V_0) & g_{2,2}(V_0) \end{pmatrix}.$$

4.1. Proof of Lemma 1

For simplicity, we prove the lemma only for the case where $m = 2$ and $k = 2$. For $t = 1, 2, \dots, n$, $i = 1, 2$ and $j = 1, 2$, let

$$\hat{\eta}_{j,1}(z_i) = (nh^d)^{-1} \sum_{t=1}^{n} (g_j^*(Z_t) - g_j^*(z_i)) k_0\left(\frac{Z_t - z_i}{h}\right),$$

$$\hat{\eta}_{j,2}(z_i) = (nh^d)^{-1} \sum_{t=1}^{n} u_{j,t} k_0\left(\frac{Z_t - z_i}{h}\right)$$

and $\hat{\eta}_j(z_i) = \hat{\eta}_{j,1}(z_i) + \hat{\eta}_{j,2}(z_i)$. Then, $\hat{g}_j(z_i) - g_j^*(z_i) = \hat{\eta}_j(z_i)/\hat{f}_Z(z_i)$, where $\hat{f}_Z(z_i) = (1/(nh^d)) \sum_{t=1}^n k_0((Z_t - z_i)/h)$. We can complete the proof using the following results (A1)–(A3).

(A1) Suppose that the conditions in Lemma 1 hold. Then, for $1 \le i, j \le 2$,

$$\hat{\eta}_{j,1}(z_i) = h^2 \sum_{s=1}^{d} B_{s,j}(z_i) + o_p\left(h^3 + (nh^d)^{-1/2}\right).$$

(A2) Suppose that the conditions in Lemma 1 hold. Then,

$$Z_n^* \equiv \sqrt{nh^d} \begin{pmatrix} \hat{\eta}_{1,2}(z_1) \\ \hat{\eta}_{2,2}(z_1) \\ \hat{\eta}_{1,2}(z_2) \\ \hat{\eta}_{2,2}(z_2) \end{pmatrix} \xrightarrow{D} Z,$$

where the distribution of Z is $N(0, \Sigma)$ and Σ is

$$\begin{pmatrix} \kappa^d \sigma_1^2(z_1) f_Z(z_1) & \kappa^d c_{12}(z_1) f_Z(z_1) & 0 & 0 \\ \kappa^d c_{12}(z_1) f_Z(z_1) & \kappa^d \sigma_2^2(z_1) f_Z(z_1) & 0 & 0 \\ 0 & 0 & \kappa^d \sigma_1^2(z_2) f_Z(z_2) & \kappa^d c_{12}(z_2) f_Z(z_2) \\ 0 & 0 & \kappa^d c_{12}(z_2) f_Z(z_2) & \kappa^d \sigma_2^2(z_2) f_Z(z_2) \end{pmatrix}.$$

(A3) Suppose that $(X_{n1}, X_{n2}, \dots, X_{nk})^T \xrightarrow{D} (Y_1, Y_2, \dots, Y_k)^T$ and $(Z_{n1}, Z_{n2} \dots, Z_{nk})^T \xrightarrow{D} (c_1, c_2, \dots, c_k)^T$, where c_1, $c_2, \dots, c_k$ are constants. Then,

$$(X_{n1}Z_{n1}, X_{n2}Z_{n2}, \dots, X_{nk}Z_{nk})^T \xrightarrow{D} (c_1Y_1, c_2Y_2, \dots, c_kY_k)^T.$$

From (A1), (A2) and the assumption that $nh^{d+4} \to 0$, we have

$$\sqrt{nh^d}\begin{pmatrix} \hat{\eta}_1(z_1) - h^2 \sum_{s=1}^{d} B_{s,1}(z_1) \\ \hat{\eta}_2(z_1) - h^2 \sum_{s=1}^{d} B_{s,2}(z_1) \\ \hat{\eta}_1(z_2) - h^2 \sum_{s=1}^{d} B_{s,1}(z_2) \\ \hat{\eta}_2(z_2) - h^2 \sum_{s=1}^{d} B_{s,2}(z_2) \end{pmatrix} \sim Z + (nh^d)^{1/2} o_p\left(h^3 + \frac{1}{\sqrt{nh^d}}\right) \xrightarrow{D} Z,$$

where $A \sim B$ means that the distributions of A and B are the same. Apply (A3) and we have Lemma 1.

The proofs of (A1)–(A3) are given below.

- Proof of (A1). Note that

$$\begin{aligned} E\left(\hat{\eta}_{j,1}(z_i)\right) &= \frac{1}{h^d}\int \left(g_j^*(z_t) - g_j^*(z_i)\right) k_0\left(\frac{z_t - z_i}{h}\right) f_Z(z_t)dz_t \\ &= \int \left(g_j^*(z_i + h\nu) - g_j^*(z_i)\right) k_0(\nu) f_Z(z_i + h\nu)d\nu \\ &\overset{\nu=(\nu_1,\dots,\nu_d)}{=} \int h \sum_{s=1}^{d} g_{j,s}^*(z_i)\nu_s f_Z(z_i)k_0(\nu)d\nu + \int h^2 \left(\sum_{s=1}^{d} g_{j,s}^*(z_i)\nu_s\right)\left(\sum_{s=1}^{d} f_s(z_i)\nu_s\right) k_0(\nu)d\nu \\ &\quad + \frac{1}{2}\int h^2 f_Z(z_i) \sum_{s=1}^{d}\sum_{s^*=1}^{d} g_{j,ss^*}^*(z_i)\nu_s\nu_{s^*}k_0(\nu)d\nu + O(h^3) \\ &= h^2\frac{\kappa_2}{2}\sum_{s=1}^{d}\left(f_Z(z_i)g_{j,ss}^*(z_i) + 2f_s(z_i)g_{j,s}^*(z_i)\right) + O(h^3) \\ &= h^2\sum_{s=1}^{d} B_{s,j}(z_i) + O(h^3). \end{aligned}$$

Let $K_{i,j,t} = h^{-d}\left(g_j^*(Z_t) - g_j^*(z_i)\right) k_0((Z_t - z_i)/h)$. Then, we have

$$\mathrm{Var}\left(\hat{\eta}_{j,1}(z_i)\right) = \frac{1}{n^2}\left(\sum_{t=1}^{n}\mathrm{Var}(K_{i,j,t}) + \sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\mathrm{Cov}(K_{i,j,t}, K_{i,j,s})\right).$$

Since

$$\begin{aligned} \mathrm{Var}(K_{i,j,t}) &= E\left(K_{i,j,t}^2\right) - \left(E(K_{i,j,t})\right)^2 \\ &= \frac{1}{h^d}\int \left(g_j^*(z_i + h\nu) - g_j^*(z_i)\right)^2 (k_0(\nu))^2 f_Z(z_i + h\nu)d\nu \\ &\quad - \left(f_Z(z_i)h\int \sum_{s=1}^{d} g_{j,s}^*(z_i)\nu_s k_0(\nu)d\nu + O(h^2)\right)^2 \\ &= \frac{1}{h^d}O(h^2) - O(h^4), \end{aligned}$$

$\sum_{t=1}^{n}\mathrm{Var}(K_{i,j,t}) = O(nh^{2-d})$. Note that from Corollary A.2 in [6] and the fact that for $2 < \beta < 2(2+d)/d$, $E(\left|K_{i,j,t}^{\beta}\right|) = O(h^{2+d-\beta d})$, we have

$$\left|\sum_{s\neq t}\mathrm{Cov}(K_{i,j,t},K_{i,j,s})\right| = \left|2\sum_{t=1}^{n}\sum_{s>t}^{n}\mathrm{Cov}(K_{i,j,t},K_{i,j,s})\right|$$
$$\le 2n\sum_{s=1}^{\infty}\left|\mathrm{Cov}(K_{i,j,1},K_{i,j,1+s})\right|$$
$$\le 16nO(h^{2(2+d-\beta d)/\beta})\sum_{s=1}^{\infty}\alpha^{(\beta-2)/\beta}(s).$$

Therefore,

$$\mathrm{Var}\left(\hat{\eta}_{j,1}(z_i)\right) = O\left(\frac{h^2}{nh^d}\right) + O\left(\frac{h^{2(2+d-\beta d)/\beta}}{n}\right) = o\left(\frac{1}{nh^d}\right).$$

From the above results, $\hat{\eta}_{j,1}(z_i) = h^2\sum_{s=1}^{d}B_{s,j}(z_i) + o_p(h^3 + (nh^d)^{(-1/2)})$.

- Proof of (A2). By the Cramér–Wold Theorem, it is sufficient to prove that $c^TZ_n^*$ converges in distribution to c^TZ for any $c = (c_1, c_2, c_3, c_4)^T$ in R^4. We use "big–small block" arguments in [2,12] to complete the proof. Assume that there exist positive integers $p = p(n)$, $q = q(n)$ and $k = k(n) = [n/(p + q)]$ (the integer part of $n/(p + q)$) such that as $n \to \infty$,

$$p \to \infty, \qquad q \to \infty, \qquad p = o(n), \qquad q = o(p), \qquad p = o\left((nh^d)^{1/2}\right),$$
$$np^{-1}\alpha(q) = o(1), \qquad ph^d = o(1), \qquad p^{\epsilon}h^d \to \infty.$$

Let

$$Z_{n,t} = \frac{1}{\sqrt{h^d}}\left(c_1u_{1,t}k_0\left(\frac{Z_t - z_1}{h}\right) + c_2u_{2,t}k_0\left(\frac{Z_t - z_1}{h}\right) + c_3u_{1,t}k_0\left(\frac{Z_t - z_2}{h}\right) + c_4u_{2,t}k_0\left(\frac{Z_t - z_2}{h}\right)\right).$$

Then, we have $c^TZ_n^* = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_{n,t} \equiv \frac{1}{\sqrt{n}}W_n$. Let $\xi_j = \sum_{t=j(p+q)+1}^{(j+1)p+jq}Z_{n,t}$ and $\zeta_j = \sum_{t=(j+1)p+jq+1}^{(j+1)(p+q)}Z_{n,t}$ for $j = 0, 1, \ldots, k-1$, and $\zeta_k = \sum_{t=k(p+q)+1}^{n}Z_{n,t}$. Then, $W_n = \underbrace{\sum_{j=0}^{k-1}\xi_j}_{W_{n1}} + \underbrace{\sum_{j=0}^{k-1}\zeta_j}_{W_{n2}} + \zeta_k$. In order to prove this lemma, it suffices to show that as $n \to \infty$,

(1) $\left|E(\exp(itW_{n1})) - \prod_{j=0}^{k-1}E(\exp(it\xi_j))\right| \to 0$,
(2) $\frac{1}{\sqrt{n}}W_{n2} \xrightarrow{p} 0$ and $\frac{1}{\sqrt{n}}\zeta_k \xrightarrow{p} 0$,
(3) $\sigma_n^2 \equiv \sum_{j=0}^{k-1}E(\xi_j^2) = n(\sigma^2 + o(1))$,
(4) $\frac{1}{\sigma_n^2}\sum_{j=0}^{k-1}E(\xi_j^2I(|\xi_j| > \varepsilon\sqrt{\sigma_n^2})) \to 0$ for any $\varepsilon > 0$,

where

$$\sigma^2 = c_1^2\kappa^df_Z(z_1)\sigma_1^2(z_1) + c_2^2\kappa^df_Z(z_1)\sigma_2^2(z_1) + c_3^2\kappa^df_Z(z_2)\sigma_1^2(z_2) + c_4^2\kappa^df_Z(z_2)\sigma_2^2(z_2)$$
$$+ 2c_1c_2\kappa^df_Z(z_1)c_{12}(z_1) + 2c_3c_4\kappa^df_Z(z_2)c_{12}(z_2).$$

The verification of the above expression for σ_n^2 is given in Section 4.5.
We now prove these results respectively. From Lemma 18.2 in [10], which is due to Volkonskii and Rozanov [16],

$$\left|E(\exp(itW_{n1})) - \prod_{j=0}^{k-1}E(\exp(it\xi_j))\right| \le 16k\alpha(q) = O\left(\frac{n}{p}\alpha(q)\right) = o(1),$$

we obtain (1). In order to prove (2), we first consider W_{n2}. Note that

$$E(W_{n2}^2) = \mathrm{Var}\left(\sum_{j=0}^{k-1}\zeta_j\right) = \underbrace{k\mathrm{Var}(\zeta_0)}_{\text{(P1)}} + \underbrace{\sum_{i=0}^{k-1}\sum_{j=0,j\neq i}^{k-1}\mathrm{Cov}(\zeta_i,\zeta_j)}_{\text{(P2)}}.$$

Computation of (P1). Note that from

$$\mathrm{Var}(\zeta_0) = \sum_{i=1}^{q}\mathrm{Var}(Z_{n,i}) + 2\sum_{i=1}^{q}\sum_{j>i}^{q}\mathrm{Cov}(Z_{n,i},Z_{n,j}),$$

$$\sum_{i=1}^{q}\mathrm{Var}(Z_{n,i}) = q\sigma^2 + O(qh^2),$$

and the fact that

$$2\sum_{i=1}^{q}\sum_{j>i}^{q}\mathrm{Cov}(Z_{n,i},Z_{n,j})=2q\sum_{j=1}^{q}\left(1-\frac{j}{q}\right)\mathrm{Cov}(Z_{n,1},Z_{n,1+j})=O(q^2h^d),$$

we have that

$$\mathrm{Var}(\zeta_0)=q\sigma^2+O(q^2h^d)+O(qh^2)=q\sigma^2(1+o(1)).$$

Therefore,

$$(\mathrm{P1})=kq\sigma^2(1+o(1))=O(kq)=o(n).$$

Computation of (P2). Note that from Theorem A.5 in [6],

$$\begin{aligned}|(\mathrm{P2})|&=\left|2\sum_{i=0}^{k-1}\sum_{j>i}^{k-1}\mathrm{Cov}(\zeta_i,\zeta_j)\right|\\&\le 2\sum_{i=1}^{n-p}\sum_{j=i+p}^{n}\left|\mathrm{Cov}(Z_{n,i},Z_{n,j})\right|\le 2n\sum_{j=p}^{\infty}\left|\mathrm{Cov}(Z_{n,1},Z_{n,1+j})\right|\\&\le 2n\sum_{j=p}^{\infty}4C_{1n}C_{2n}\alpha(j)\le C^*\frac{n}{h^d}\sum_{j=p}^{\infty}\alpha(j)=o(n),\end{aligned}$$

where $C_{in}=4\max|c_k|\sup|u_{s,1}|\sup|k_0|/\sqrt{h^d}$ for $i=1,2$. Then, we have $E(W_{n2}^2)/n=o(1)$. Similarly, $\mathrm{Var}(\zeta_k)=O(p+q)=o(n)$, so (2) holds.

By stationarity and the same arguments in (1), we have $\mathrm{Var}(\xi_0)=p\sigma^2(1+o(1))$. Thus $\sum_{j=0}^{k-1}E(\xi_j^2)/n=kp\sigma^2(1+o(1))/n\to\sigma^2$. Finally, since $\left|Z_{n,t}\right|\le C/\sqrt{h^d}$, for every $\epsilon>0$, the set $\{|\xi_j|\ge\epsilon\sqrt{\sigma_n^2}\}$ is an empty set when n is large. Therefore, (4) holds. This completes the proof.

- Proof of (A3). It is sufficient to prove that $(X_{n1},\dots,X_{nk},Z_{n1},\dots,Z_{nk})^T\xrightarrow{D}(Y_1,\dots,Y_k,c_1,\dots,c_k)$. Let $X_n=(X_{n1},\dots,X_{nk})^T$, $Z_n=(Z_{n1},\dots,Z_{nk})^T$, $Y=(Y_1,\dots,Y_k)^T$ and $c=(c_1,\dots,c_k)^T$. Then,

$$\begin{aligned}E(e^{i(t^TX_n+s^TZ_n)})&=E(e^{i(t^TX_n+s^Tc)}e^{i(s^T(Z_n-c))})\\&=\underbrace{E(e^{i(t^TX_n+s^Tc)}(e^{i(s^T(Z_n-c))}-1))}_{I}+\underbrace{E(e^{i(t^TX_n+s^Tc)})}_{II}.\end{aligned}$$

Note that $II\to E(e^{i(t^TY+s^Tc)})$ and $I\to 0$ by Lebesgue's dominated convergence theorem. Apply the continuous mapping theorem and we have (A3).

4.2. Proof of Theorem 1

We adopt the proof in [8]. For $z\in\{z_1,\dots,z_k\}$, let ϕ_l^*: $1\le l\le p$ and $\psi_{m^*}^*:1\le m^*\le q$ be the new basis functions obtained by making linear transformations of ϕ_l's and ψ_{m^*}'s such that $\phi_1^*=1=\psi_1^*$, $(E(\phi_l^*(X)\phi_{l'}^*(X)|Z=z):1\le l,l'\le p)$ and $(E(\psi_{m^*}^*(Y)\psi_{m'}^*(Y)|Z=z):1\le m^*,m'\le q)$ are identity matrices, and $E(\phi_l^*(X)\psi_{m^*}^*(Y)|Z=z)=0$ for $l\ne m^*$. Take $g_1(X,Y),\dots,g_m(X,Y)$ to be the functions $\phi_l^*(X)\phi_{l'}^*(X)$, $\phi_l^*(X)\psi_{m^*}^*(Y)$ and $\psi_{m^*}^*(Y)\psi_{m'}^*(Y)$, where $1\le l\le l'\le p$ and $1\le m^*\le m'\le q$. Apply Lemma 1 and we have

$$\sqrt{nh^d}\begin{pmatrix}\hat g_1(z_1)-g_1^*(z_1)\\\vdots\\\hat g_1(z_k)-g_1^*(z_k)\\\vdots\\\hat g_m(z_1)-g_m^*(z_1)\\\vdots\\\hat g_m(z_k)-g_m^*(z_k)\end{pmatrix}-\sqrt{nh^d}\begin{pmatrix}h^2\sum_{s=1}^{d}B_{s,1}(z_1)/f_Z(z_1)\\\vdots\\h^2\sum_{s=1}^{d}B_{s,1}(z_k)/f_Z(z_k)\\\vdots\\h^2\sum_{s=1}^{d}B_{s,m}(z_1)/f_Z(z_1)\\\vdots\\h^2\sum_{s=1}^{d}B_{s,m}(z_k)/f_Z(z_k)\end{pmatrix}\xrightarrow{D}Z^*. \qquad (18)$$

Let

$$V^*(z) = \begin{pmatrix} V_{11}(z) & V_{12}(z) \\ V_{21}(z) & V_{22}(z) \end{pmatrix},$$

where the (l, l')-th element of $V_{11}(z)$ is $E(\phi_l^*(X)\phi_{l'}^*(X)|Z = z)$ for $1 \le l, l' \le p$, the (l, m^*)-th element of $V_{12}(z)$ is $E(\phi_l^*(X)\psi_{m^*}^*(Y)|Z = z)$ for $1 \le l \le p$, $1 \le m^* \le q$, the (m^*, m')-th element of $V_{22}(z)$ is $E(\psi_{m^*}^*(Y)\psi_{m'}^*(Y)|Z = z)$ for $1 \le m^*, m' \le q$, and $V_{21}(z) = (V_{12}(z))^T$. Let $\hat{V}^*(z)$ be the estimator of $V^*(z)$ obtained by replacing each conditional expectation in $V^*(z)$ with its kernel estimator defined in (8). Then, (18) gives

$$\sum_{i=1}^{k} \|\hat{V}^*(z_i) - V^*(z_i)\|^2 = O_p\left(\frac{1}{nh^d}\right) + O_p(h^4) = O_p\left(\frac{1}{nh^d} + h^4\right).$$

For $1 \le i \le k$, for a $p \times 1$ vector a and a $(p+q) \times (p+q)$ matrix

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix},$$

wherethe dimensions of U_{11}, U_{12}, U_{21} and U_{22} are $p \times p$, $p \times q$, $q \times p$ and $q \times q$ respectively, define

$$g_{r,s}(U) = U_{rs} \tag{19}$$

for $1 \le r, s \le 2$,

$$g_{r,s}^*(U) = \begin{cases} g_{r,s}(U) & \text{if } (r, s) = (1, 2) \text{ or } (2, 1); \\ (g_{r,s}(U))^{-1} & \text{if } (r, s) = (1, 1) \text{ or } (2, 2), \end{cases}$$

and

$$g(U, a) = U_{1,2}U_{2,2}^{-1}U_{2,1}U_{1,1}^{-1} - U_{1,1}aa^T. \tag{20}$$

Let α^* be the $p \times 1$ vector whose first element is 1 and the rest elements are 0's. Then, $\hat{\rho}(z)$ and $\rho_{p,q}(z)$ are the square roots of the largest eigenvalues of the matrices $g(\hat{V}^*(z), \alpha^*)$ and $g(V^*(z), \alpha^*)$ respectively. Let

$$\triangle_{r,s,i} = g_{r,s}^*(\hat{V}^*(z_i)) - g_{r,s}^*(V^*(z_i)).$$

Then, we have

$$\begin{aligned} \| g(\hat{V}^*(z_i), \alpha^*) - g(V^*(z_i), \alpha^*)\| \le & \prod_{r=1}^{2}\prod_{s=1}^{2}(\|g_{r,s}^*(V^*(z_i))\| + \| \triangle_{r,s,i} \|) - \prod_{r=1}^{2}\prod_{s=1}^{2}\|g_{r,s}^*(V^*(z_i))\| \\ & + \|g_{1,1}^*(\hat{V}^*(z_i)) - g_{1,1}^*(z_i)\| \, \|\alpha^*(\alpha^*)^T\|, \end{aligned}$$

which gives that

$$\sum_{i=1}^{k} \|g(\hat{V}^*(z_i), \alpha^*) - g(V^*(z_i), \alpha^*)\|^2 = O_p\left(\frac{1}{nh^d} + h^4\right) = O_p\left(\frac{1}{nh^d}\right)$$

and

$$\sum_{i=1}^{k} \left(\hat{\rho}^2(z_i) - \rho_{p,q}^2(z_i)\right)^2 = O_p\left(\frac{1}{nh^d} + h^4\right) \tag{21}$$

since $|\hat{\rho}^2(z_i) - \rho_{p,q}^2(z_i)| \le \|g(\hat{V}^*(z_i), \alpha^*) - g(V^*(z_i), \alpha^*)\|$ for $1 \le i \le k$. From (21) and the fact that

$$\sum_{i=1}^{k}(\hat{f}_Z(z_i) - f_Z(z_i))^2 = O_p\left(\frac{1}{nh^d} + h^4\right),$$

$$\begin{aligned} \left(\sum_{i=1}^{k}\hat{f}_Z(z_i)\hat{\rho}^2(z_i) - \sum_{i=1}^{k}f_Z(z_i)\rho_{p,q}^2(z_i)\right)^2 &= \left(\sum_{i=1}^{k}(\hat{f}_Z(z_i) - f_Z(z_i))\hat{\rho}^2(z_i) + \sum_{i=1}^{k}f_Z(z_i)(\hat{\rho}^2(z_i) - \rho_{p,q}^2(z_i))\right)^2 \\ &= O_p\left(\frac{1}{nh^d} + h^4\right). \end{aligned}$$

4.3. Proof of Theorem 2

We adopt the proof in [8]. For $z \in \{z_1, \ldots, z_k\}$, let $\hat{V}^*(z)$, $V^*(z)$ and $B_{s,j}$ be as defined in the proof of Theorem 1. Let $\mathbf{B}_i$ be the $(p+q)\times(p+q)$ matrix whose elements are $h^2\sum_{s=1}^d B_{s,j}(z_i)/f_Z(z_i)$: $1 \le j \le m = (p+q)^2$. From Lemma 1, we have

$$\begin{pmatrix} \sqrt{nh^d f_Z(z_1)/\kappa^d}(\hat{V}^*(z_1) - V^*(z_1) - \mathbf{B}_1) \\ \vdots \\ \sqrt{nh^d f_Z(z_k)/\kappa^d}(\hat{V}^*(z_k) - V^*(z_k) - \mathbf{B}_k) \end{pmatrix} \xrightarrow{D} \begin{pmatrix} N_1^* \\ \vdots \\ N_k^* \end{pmatrix} \equiv N^*,$$

where for $1 \le i \le k$, N_i^* is a normal matrix of elements with mean 0 and variance 1. Apply the Skorohod's theorem, for $1 \le i \le k$, there exist random matrices T_i and $W_{1,i}$ such that $T_i \sim (nh^d f_Z(z_i)/\kappa^d)^{1/2}(\hat{V}^*(z_i) - V^*(z_i) - \mathbf{B}_i)$, $W_{1,i} \sim N_i^*$ and $T_i \to W_{1,i}$ almost surely. Therefore,

$$\hat{V}^*(z_i) \sim \frac{\sqrt{\kappa^d}T_i}{\sqrt{nh^d f_Z(z_i)}} + V^*(z_i) + \mathbf{B}_i = V^*(z_i) + \frac{\sqrt{\kappa^d}}{\sqrt{nh^d f_Z(z_i)}}(W_{1,i} + W_{2,i}),$$

where $W_{2,i} = T_i - W_{1,i} + \sqrt{nh^d f_Z(z_i)/\kappa^d}\mathbf{B}_i$. Note that $\mathbf{B}_i = O(h^2)$. From (S6), $\sum_{i=1}^k \|W_{2,i}\| = o_p(1)$.

For $1 \le i \le k$, let $\tilde{V}_i = V^*(z_i) + (nh^d f_Z(z_i)/\kappa^d)^{-1/2}(W_{1,i} + W_{2,i})$, $A_1(z_i) = g(\tilde{V}_i, \alpha^*)g_{1,1}(\tilde{V}_i)$ and $\tilde{\rho}_0^2(z_i)$ be the largest eigenvalue of $A_1(z_i)(g_{1,1}(\tilde{V}_i))^{-1}$. Here the functions $g(\cdot,\cdot)$ and $g_{1,1}$ are defined in (20) and (19) respectively. Then, $\tilde{\rho}_0(z_i)$ has the same distribution as $\hat{\rho}(z_i)$. Below we will show that the impact of $W_{2,i}$ is negligible in the derivation of the asymptotic distribution of $\tilde{\rho}_0(z_i)$.

For $1 \le r, s \le 2$ and $1 \le i \le k$, let $\triangle_{r,s,i} = g_{r,s}(\tilde{V}_i) - g_{r,s}(V^*(z_i))$. Then,

$$\sum_{i=1}^k \sum_{r=1}^2 \sum_{s=1}^2 \| \triangle_{r,s,i} \|^2 = O_p\left(\frac{1}{nh^d} + h^4\right) = O_p\left(\frac{1}{nh^d}\right)$$

and

$$\begin{aligned} A_1(z_i) = {} & g_{1,2}(V^*(z_i))(g_{2,2}(\tilde{V}_i))^{-1}g_{2,1}(V^*(z_i)) - g_{1,1}(\tilde{V}_i)\alpha^*(\alpha^*)^T g_{1,1}(\tilde{V}_i) \\ & + g_{1,2}(V^*(z_i))\triangle_{2,1,i} + \triangle_{1,2,i}\, g_{2,1}(V^*(z_i)) + \triangle_{1,2,i}\triangle_{2,1,i} \\ & - g_{1,2}(V^*(z_i))\triangle_{2,2,i}\triangle_{2,1,i} - \triangle_{1,2,i}\triangle_{2,2,i}\, g_{2,1}(V^*(z_i)) + R_{1,i}, \end{aligned}$$

where

$$\begin{aligned} R_{1,i} = {} & \triangle_{1,2,i}((g_{2,2}(\tilde{V}_k))^{-1} - I_q)\triangle_{2,1,i} + g_{1,2}(V^*(z_i))((g_{2,2}(\tilde{V}_i))^{-1} - I_q + \triangle_{2,2,i})\triangle_{2,1,i} \\ & + \triangle_{1,2,i}((g_{2,2}(\tilde{V}_i))^{-1} - I_q + \triangle_{2,2,i})g_{2,1}(V^*(z_i)) \end{aligned}$$

and I_q denotes the $q \times q$ identity matrix. Note that $g_{2,2}(\tilde{V}_i)$ can be expressed as

$$g_{2,2}(\tilde{V}_i) = \begin{pmatrix} 1 & B_i^T \\ B_i & D_i \end{pmatrix}$$

for some matrices B_i and D_i, so $A_1(z_i)$ becomes

$$\begin{aligned} & B_i^T((D_i - B_iB_i^T)^{-1} - I_{q-1})B_iJ + g_{1,2}(V^*(z_i))(\triangle_{2,2,i} - J)^2 g_{2,1}(V^*(z_i)) \\ & - \triangle_{1,1,i}\, g_{1,2}(V^*(z_i))g_{2,1}(V^*(z_i))\triangle_{1,1,i} + \triangle_{1,2,i}\triangle_{2,1,i} \\ & - g_{1,2}(V^*(z_i))\triangle_{2,2,i}\triangle_{2,1,i} - \triangle_{1,2,i}\triangle_{2,2,i}\, g_{2,1}(V^*(z_i)) + R_{1,i}, \end{aligned}$$

where $J = \alpha^*(\alpha^*)^T$. Let

$$\begin{aligned} A_2(z_i) = {} & g_{1,2}(V^*(z_i))(g_{2,2}(W_{1,i}))^2 g_{2,1}(V^*(z_i)) \\ & - g_{1,1}(W_{1,i})g_{1,2}(V^*(z_i))g_{2,1}(V^*(z_i))g_{1,1}(W_{1,i}) + g_{1,2}(W_{1,i})g_{2,1}(W_{1,i}) \\ & - g_{1,2}(V^*(z_i))g_{2,2}(W_{1,i})g_{2,1}(W_{1,i}) - g_{1,2}(W_{1,i})g_{2,2}(W_{1,i})g_{2,1}(V^*(z_i)) \end{aligned}$$

and

$$\begin{aligned} R_{2,i} = {} & B_i^T((D_i - B_iB_i^T)^{-1} - I_{q-1})B_iJ \\ & - (nh^d f_Z(z_i)/\kappa^d)^{-1}A_2(z_i) + g_{1,2}(V^*(z_i))(\triangle_{2,2,i} - J)^2 g_{2,1}(V^*(z_i)) \\ & - \triangle_{1,1,i}\, g_{1,2}(V^*(z_i))g_{2,1}(V^*(z_i))\triangle_{1,1,i} + \triangle_{1,2,i}\triangle_{2,1,i} \\ & - g_{1,2}(V^*(z_i))\triangle_{2,2,i}\triangle_{2,1,i} - \triangle_{2,1,i}\triangle_{2,2,i}\, g_{2,1}(V^*(z_i)). \end{aligned}$$

Then,

$$A_1(z_i) = \frac{A_2(z_i)\kappa^d}{nh^d f_Z(z_i)} + R_{1,i} + R_{2,i}, \tag{22}$$

where

$$\sum_{i=1}^{k}(\|R_{1,i}\|^2 + \|R_{2,i}\|^2) = O_p\left(\frac{1}{(nh^d)^2}\right). \tag{23}$$

Note that under conditional independence, for $1 \le i \le k$, $A_2(z_i) = C_i C_i^T$, where C_i is the $p \times q$ matrix obtained by replacing elements in the first column and first row of $g_{1,2}(W_{1,i})$ with zero's, and $g_{1,2}(W_{1,i})$ is a random matrix whose elements are IID $N(0, 1)$ expect that the $(1, 1)$-th element is 1. Therefore, $\sum_{i=1}^{k} \|A_2(z_i)\|^2 = O_p(1)$, which, together with (22) and (23), implies that $\sum_{i=1}^{k} \|A_1(z_i)\|^2 = O_p(1/((nh^d)^2))$ and

$$\sum_{i=1}^{k} \|A_1(z_i)(g_{1,1}(\tilde{V}_i))^{-1} - A_1(z_i)\|^2 = O_p\left(\frac{1}{(nh^d)^3}\right). \tag{24}$$

For $1 \le i \le k$, let $\lambda_{0,i}$ be the largest eigenvalue of $A_2(z_i)$. By (22)–(24),

$$\sum_{i=1}^{k}(nh^d f_Z(z_i)\tilde{\rho}_0^2(z_i)/\kappa^d - \lambda_{0,i})^2 = o_p(1).$$

Let $\tilde{f}_i$, $\tilde{\rho}(z_i)$ and $\lambda_i : 1 \le i \le k$ be random variables such that the joint distribution of $(\tilde{f}_i, \tilde{\rho}(z_i)) : 1 \le i \le k$ is the same as $(\hat{f}_Z(z_i), \hat{\rho}(z_i)) : 1 \le i \le k$, and the joint distribution of $(\tilde{\rho}(z_i), \lambda_i) : 1 \le i \le k$ is the same as $(\tilde{\rho}_0(z_i), \lambda_{0,i}) : 1 \le i \le k$. Note that $nh^d \sum_{i=1}^{k}(\hat{\rho}(z_i))^2 = O_p(1)$, so we have that

$$\begin{aligned}\left|\frac{nh^d}{\kappa^d}\sum_{i=1}^{k}\hat{f}_Z(z_i)(\hat{\rho}(z_i))^2 - \frac{nh^d}{\kappa^d}\sum_{i=1}^{k}f_Z(z_i)(\hat{\rho}(z_i))^2\right| &\le \frac{nh^d}{\kappa^d}\left(\sum_{i=1}^{k}(\hat{f}_Z(z_i) - f_Z(z_i))^2\right)^{1/2}\sum_{i=1}^{k}(\hat{\rho}(z_i))^2 \\ &= O_p(1)O_p((nh^d)^{-1/2}) = O_p((nh^d)^{-1/2})\end{aligned}$$

and

$$\left|\frac{nh^d}{\kappa^d}\sum_{i=1}^{k}\tilde{f}_i(\tilde{\rho}(z_i))^2 - \sum_{i=1}^{k}\lambda_i\right| \le O_p((nh^d)^{-1/2}) + o_p(1) = o_p(1).$$

The proof of Theorem 2 is complete.

4.4. *Proof of Theorem 3*

Suppose that $\rho(z_i) > 0$ for some z_i. Then, we have $\sum_{i=1}^{k} f_Z(z_i)\rho^2(z_i) > 0$. Choose ϵ such that $0 < \epsilon < \sum_{i=1}^{k} f_Z(z_i)\rho^2(z_i)$ and we have

$$\underbrace{P\left(\sum_{i=1}^{k}\hat{f}_Z(z_i)\hat{\rho}^2(z_i) \ge \sum_{i=1}^{k}f_Z(z_i)\rho^2(z_i) - \epsilon\right)}_{\text{III}} \le P\left(\sum_{i=1}^{k}\hat{f}_Z(z_i)\hat{\rho}^2(z_i) \ge \frac{\kappa^d F^*_{1-\alpha}}{nh^d}\right)$$

for large n. From Theorem 1,

$$\text{III} \ge P\left(\left|\sum_{i=1}^{k}\hat{f}_Z(z_i)\hat{\rho}^2(z_i) - \sum_{i=1}^{k}f_Z(z_i)\rho^2(z_i)\right| \le \epsilon\right) \to 1,$$

so

$$P\left(\sum_{i=1}^{k}\hat{f}_Z(z_i)\hat{\rho}^2(z_i) \ge \frac{\kappa^d F^*_{1-\alpha}}{nh^d}\right) \to 1.$$

4.5. The verification of the expression for σ_n^2

The expression for σ_n^2 involves some variance and covariance terms. Under the conditions in Theorem 1, the major parts for those variance and covariance terms can be obtained. The results are as follows. For $1 \le i, i^* \le k$ and $1 \le j, j^* \le m$, 1–4 hold.

1. $\mathrm{Var}\left(u_{j,t}k_0\left(\frac{Z_t-z_i}{h}\right)\right) = h^d\kappa^d\sigma_j^2(z_i)f_Z(z_i) + O(h^{d+2})$.
2. $\mathrm{Cov}\left(u_{j,t}k_0\left(\frac{Z_t-z_i}{h}\right), u_{j^*,t}k_0\left(\frac{Z_t-z_i}{h}\right)\right) = h^d\kappa^d c_{jj^*}(z_i)f_Z(z_i) + O(h^{d+2})$.
3. $\mathrm{Cov}\left(u_{j,t}k_0\left(\frac{Z_t-z_i}{h}\right), u_{j,t}k_0\left(\frac{Z_t-z_{i^*}}{h}\right)\right) = O(h^{2d})$.
4. $\mathrm{Cov}\left(u_{j,t}k_0\left(\frac{Z_t-z_i}{h}\right), u_{j^*,t}k_0\left(\frac{Z_t-z_{i^*}}{h}\right)\right) = O(h^{2d})$.

We will only give the proof for Case 1 since the proofs for other cases are similar. Since

$$\begin{aligned}
\mathrm{Var}\left(u_{j,t}k_0\left(\frac{Z_t-z_i}{h}\right)\right) &= E\left(E\left(u_{j,t}^2\left(k_0\left(\frac{Z_t-z_i}{h}\right)\right)^2|Z_t\right)\right) \\
&= \int \sigma_j^2(z_t)\left(k_0\left(\frac{z_t-z_i}{h}\right)\right)^2 f_Z(z_t)dz_t \\
&= h^d\int \sigma_j^2(z_i+h\nu)(k_0(\nu))^2 f_Z(z_i+h\nu)d\nu \\
&= h^d\int \sigma_j^2(z_i)(k_0(\nu))^2\left(f_Z(z_i)+h\sum_{s=1}^d f_s(z_i)\nu_s+O(h^2)\right)d\nu \\
&= h^d\kappa^d\sigma_j^2(z_i)f_Z(z_i)+O(h^{d+2}),
\end{aligned}$$

we complete verification of the expression for σ_n^2.

Acknowledgments

This research is supported by National Science Council in Taiwan under grant NSC 99-2118-M-004-006. The authors would like to thank the reviewers for careful reading and constructive comments.

References

[1] Taoufik Bouezmarni, Jeroen V.K. Rombouts, Abderrahim Taamouti, A nonparametric copula based test for conditional independence with applications to granger causality, Working papers, Departamento de Economía Universidad Carlos III de Madrid, 2009.
[2] J.V. Castellana, M.R. Leadbetter, On smoothed probability density estimation for stationary processes, Stochastic Processes and their Applications 21 (1986) 179–193.
[3] Miguel A. Delgado, Wenceslao González Manteiga, Significance testing in nonparametric regression based on the bootstrap, The Annals of Statistics 29 (5) (2001) 1469–1507.
[4] J.P. Florens, Denis Fougere, Noncausality in continuous time, Econometrica 64 (5) (1996) 1195–1212.
[5] J.P. Florens, M. Mouchart, A note on noncausality, Econometrica 50 (3) (1982) 583–591.
[6] P.G. Hall, C.C. Heyde, Martingale Limit Theory and its Applications, Academic Press, 1980.
[7] Craig Hiemstra, Jonathan D. Jones, Testing for linear and nonlinear granger causality in the stock price-volume relation, The Journal of Finance 49 (5) (1994) 1639–1664.
[8] Tzee Ming Huang, Testing conditional independence using maximal nonlinear conditional correlation, The Annals of Statistics 38 (4) (2010) 2047–2091.
[9] Lexin Li, R. Dennis Cook, Christopher J. Nachtsheim, Model-free variable selection, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (2) (2005) 285–299.
[10] Qi Li, Jeffrey Scott Racine, Nonparametric Econometrics: Theory and Practice, Princeton University Press, 2007.
[11] Oliver Bruce Linton, Pedro Gozalo, Conditional independence restrictions: testing and estimation, Cowles Foundation discussion papers 1140, Cowles Foundation, Yale University, 1996.
[12] Elias Masry, Multivariate regression estimation: local polynomial fitting for time series (corr: 97V67 p281), Stochastic Processes and their Applications 65 (1996) 81–101.
[13] Efstathios Paparoditis, Dimitris N. Politis, The local bootstrap for kernel estimators under general dependence conditions, Annals of the Institute of Statistical Mathematics 52 (1) (2000) 139–159.
[14] Liangjun Su, Halbert White, A consistent characteristic function-based test for conditional independence, Journal of Econometrics 141 (2007) 807–834.
[15] Liangjun Su, Halbert White, A nonparametric Hellinger metric test for conditional independence, Econometric Theory 24 (4) (2008) 829–864.
[16] V.A. Volkonskii, Yu.A. Rozanov, Some limit theorems for random functions. I, Theory of Probability and its Applications 4 (2) (1959) 178–197.