Probability in the Engineering and Informational Sciences, **36**, 2022, 126–143.
doi:10.1017/S026996482000039X

HIGH-ORDER CONDITIONAL DISTANCE COVARIANCE WITH CONDITIONAL MUTUAL INDEPENDENCE

PENGFEI LIU

School of Mathematics and Statistics and Research Institute of Mathematical Sciences (RIMS), Jiangsu Normal University, 101 Shanghai Road, Tongshan, Xuzhou 221116, China

Jiangsu Provincial Key Laboratory of Educational Big Data Science and Engineering, Jiangsu Normal University, 101 Shanghai Road, Tongshan, Xuzhou 221116, China

E-mail: liupengfei@jsnu.edu.cn

XUEJUN MA

School of Mathematical Sciences, Soochow University, 1 Shizi Street, Suzhou 215006, China

E-mail: xuejunma@suda.edu.cn

WANG ZHOU

Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546, Singapore

E-mail: stazw@nus.edu.sg

We construct a high-order conditional distance covariance, which generalizes the notation of conditional distance covariance. The joint conditional distance covariance is defined as a linear combination of conditional distance covariances, which can capture the joint relation of many random vectors given one vector. Furthermore, we develop a new method of conditional independence test based on the joint conditional distance covariance. Simulation results indicate that the proposed method is very effective. We also apply our method to analyze the relationships of $PM_{2.5}$ in five Chinese cities: Beijing, Tianjin, Jinan, Tangshan and Qinhuangdao by the Gaussian graphical model.

Keywords: conditional distance covariance, conditional mutual independence, local bootstrap

1. INTRODUCTION

Let $X = (X_1^\top, \ldots, X_q^\top)^\top$ be a random vector, where $X_i \in R^{p_i},\ i = 1, \ldots, q, \sum_{i=1}^q p_i = p$, Z is an r-dimensional random vector. Our goal is to test the conditional independence given a set of random variables, that is, to test the null hypothesis:

$$H_0 : X_1, \ldots, X_q \text{ are conditionally independen given } Z$$
$$\text{versus}$$
$$H_1 : \text{negation of } H_0$$

© The Author(s), 2020. Published by Cambridge University Press 0269-9648/19 $25.00

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

If H_0 is not rejected, then Z is the most important among $X_1, \ldots, X_q, Z$ because Z contains all the common part of X_i's. Otherwise, after removing Z from all X_i's, which is the same as conditioning on Z, X_i's should be dependent. In other words, we are going to find the most important variables among a set. This has many real applications. In the following, we list three real examples.

1. In the past decade, air pollution raises people's concerns in China, especially in Beijing. Many people study the index of $PM_{2.5}$ which measures the main pollutant. We collect $PM_{2.5}$ data of five Chinese cities, Beijing (BJ), Tianjin (TJ), Jinan (JN), Tangshan (TS) and Qinhuangdao (QHD), from January 1, 2017 to December 31, 2017. The data set contains 365 observations. Our aim is to analyze the regional relationships of $PM_{2.5}$ in these five Chinese cities, so that one can find the city which influences the others to a large extent. We will analyze this data set later.
2. In the medical study, people are interested in genes which are related to a particular disease. However, there are too many related genes. Hence, it is important to find the most related several genes, so that one can control these genes and cure the disease.
3. In the stock market, one may ask if the set of stocks which contribute to the calculation of the stock index is really the most important. This is just our hypothesis test problem.

If the joint distribution of all variables is normal, then one can use the Gaussian graphical model [5,8] to find the most important subset. In other words, if each variable is represented by a vertex, then this subset is the minimal set which has the following property. If one removes this set and all edges which are connected to this set, the remaining vertices in the graphical model are all isolated. Practically, one does not have joint normality. So what to do in order to find Z? A solution is offered as follows.

First of all, we assume that all variables are jointly normal. Under this extra assumption, one can find Z using the Gaussian graphical model. Secondly, we need to test the above H_0 to see if Z is really important.

As we know, testing conditional independence is an important problem in multivariate analysis. There are many papers to tackle this problem in recent years. For example, Linton and Gozalo [7] developed a nonparametric test of the hypothesis of conditional independence based on a generalization of the empirical distribution function. Su and White [9] proposed smoothed empirical likelihood ratio tests. Wang *et al.* [12] proposed the conditional distance correlation (CDC) method. We also refer to [6]. However, q is not larger than 2 in all these papers.

In this paper, we will introduce the concepts of a high-order conditional distance covariance (HCDCov) and a joint high-order conditional distance covariance (JHCDCov) to test conditional independence in the case of $q > 2$, which extends the conditional distance covariance to multivariate random variables. Now, we briefly review the distance covariance and its extensions in the literature. Distance covariance is a nonparametric method to test independence between two random vectors via the distance between the joint characteristic function and the product of marginal characteristic functions [10]. Wang *et al.* [12] developed the conditional version by considering conditional characteristic functions. Recently, Chakraborty and Zhang [3] generalized the distance covariance to a high-order distance covariance and a joint distance covariance to test independence for several random vectors.

Although HCDCov is a generalization of the idea used in [12], the technical details are far more complicated. In fact, the case of $q > 2$ is totally different from $q \leq 2$. Furthermore, compared with [3], our method is conditional, which makes the problem more complicated. Below we summarize our contributions. (1) JHCDCov not only captures the linear relations

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

among variables given others but also describes the nonlinear relationships. (2) We develop the conditional independence test in the case of $q > 2$ for the first time. Furthermore, the dimension of X_i can be larger than one, that is, $p_i \geq 1$. (3) We construct plug-in estimator methods, which are simple to use. No integration or optimization is involved.

The rest of this paper is organized as follows. In Section 2, we introduce the definition of HCDCov and JHCDCov. Asymptotic properties of their estimators are also established under some mild conditions. A new method of testing conditional independence by JHCDCov is discussed in Section 4. Section 5 assesses the finite sample performance of the proposed approach via some simulation studies. The proofs are gathered in the Appendix.

2. HCDC AND JHCDC

Let $\phi_{X_i|Z}(t_i)$ be the characteristic function of X_i given Z, that is, $\phi_{X_i|Z}(t_i) = E[\exp(\imath\langle t_i, X_i\rangle)|Z]$, where $t_i \in R^{p_i}$. Here, $\imath$ is the imaginary unit. Define $\omega_p(t) = (c_p|t|_p^{p+1})^{-1}$ with $c_p = \pi^{(1+p)/2}/\Gamma((1+p)/2)$, where $\Gamma(\cdot)$ is the gamma function, $|\cdot|_p$ is the Euclidean norm of R^p. Let $d\omega = (c_{p_1} \times \cdots \times c_{p_d} \times |t_1|_{p_1}^{p_1+1} \times \cdots \times |t_d|_{p_q}^{p_q+1})^{-1}\, dt_1 \times \cdots \times dt_q$. Denote by I_k^q the collection of k-tuples of indices from $\{1, \ldots, q\}$ such that all indices are different.

Definition 2.1: *The qth order conditional distance covariance among random vectors $X_1, \ldots, X_q$ given Z is defined as the positive square root of*

$$\mathrm{HCDCov}^2(X_1, \ldots, X_q \mid Z) = \int_{R^p} \left| E\left\{ \left[\prod_{i=1}^{q} (\phi_{X_i|Z}(t_i) - \exp(\imath\langle t_i, X_i\rangle)) \right] \middle| Z \right\} \right|^2 d\omega \quad \textbf{(2.1)}$$

HCDCov measures the difference between the joint characteristic function and its marginal, which can describe the linear and nonlinear relationships among the variables given Z. When $q = 2$, HCDCov reduces to CDCov in [12]. This follows from a direct calculation.

For the sake of presentation, given $x, x' \in R^{p_i}$, we define the conditional bivariate function

$$\begin{aligned} U_i(x, x', Z = z) = {} & E(|x - X_i'|_{p_i} \mid Z' = z) + E(|X_i - x'|_{p_i} \mid Z = z) \\ & - |x - x'|_{p_i} - E(|X_i - X_i'|_{p_i} \mid Z = z, Z' = z), \end{aligned} \quad \textbf{(2.2)}$$

where (X_i', Z') is an independent copy of (X_i, Z).

Proposition 2.1: *For $1 \leq i \leq q$,*

$$U_i(x, x', Z = z) = \int_{R^{p_i}} \left\{ [\phi_{X_i|Z=z}(t) - \exp(\imath\langle t, x\rangle)]\overline{[\phi_{X_i'|Z'=z}(t) - \exp(\imath\langle t, x'\rangle)]} \right\} \omega_{p_i}(t)\, dt$$

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

From Proposition 2.1, HCDCov can be re-expressed as

$$\begin{aligned}
&\mathrm{HCDCov}^2(\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{q}\mid \mathrm{Z}=\mathrm{z})\\
&\quad=\int_{R^p} E\left[\prod_{i=1}^{q}\left(\phi_{X_i|Z}(t_i)-\exp(\imath\langle t_i,X_i\rangle)\right)\,\middle|\, Z=z\right]\\
&\qquad\times E\left[\prod_{i=1}^{q}\overline{\left(\phi_{X_i|Z}(t_i)-\exp(\imath\langle t_i,X_i\rangle)\right)}\,\middle|\, Z=z\right]d\omega\\
&\quad=\int_{R^p} E\left[\prod_{i=1}^{q}\left(\phi_{X_i|Z}(t_i)-\exp(\imath\langle t_i,X_i\rangle)\right)\,\middle|\, Z=z\right]\\
&\qquad\times E\left[\prod_{i=1}^{q}\overline{\left(\phi_{X_i'|Z'}(t_i)-\exp(\imath\langle t_i,X_i'\rangle)\right)}\,\middle|\, Z'=z\right]d\omega\\
&\quad=E\left[\prod_{i=1}^{q}U_i(X_i,X_i',Z=z)\,\middle|\, Z=z,Z'=z\right] \qquad \textbf{(2.3)}
\end{aligned}$$

HCDCov has a shortcoming that zero HCDCov does not imply $X_1, X_2, \ldots X_q$ are independent given Z. We take $q=3$ as an example. If $X_1\perp(X_2,X_3)\mid Z$ or $X_2\perp(X_1,X_3)|Z$ or $X_3\perp(X_1,X_2)\mid Z$, then $1/3\mathrm{HCDCov}^2(X_1,X_2,X_3\mid Z)=0$. On the other hand, zero HCDCov implies

$$\begin{aligned}
&\phi_{X_1,X_2,X_3|Z}(t_1,t_2,t_3)-\phi_{X_1|Z}(t_1)\phi_{X_2|Z}(t_2)\phi_{X_3|Z}(t_3)\\
&\quad=\phi_{X_1|Z}(t_1)\phi_{X_2,X_3|Z}(t_2,t_3)+\phi_{X_2|Z}(t_2)\phi_{X_1,X_3|Z}(t_1,t_3)\\
&\qquad+\phi_{X_3|Z}(t_3)\phi_{X_1,X_2|Z}(t_1,t_2)-3\phi_{X_1|Z}(t_1)\phi_{X_2|Z}(t_2)\phi_{X_3|Z}(t_3)
\end{aligned}$$

Hence, the higher-order effect can be represented by low-order and pairwise effects. In other words, $\phi_{X_1,X_2,X_3|Z}(t_1,t_2,t_3)-\phi_{X_1|Z}(t_1)\phi_{X_2|Z}(t_2)\phi_{X_3|Z}(t_3)$ can be expressed by $\phi_{X_i|Z}(t_i)$ and $\phi_{X_j,X_k|Z}(t_j,t_k)$ for $1\le i\ne j\ne k\le 3$. If low-order effects are zero, then higher-order effect is also zero, which motives us to study JHCDCov.

DEFINITION 2.2: *The qth order conditional distance covariance among random vectors $X_1,\ldots,X_q$ given Z is defined as the positive square root of*

$$\begin{aligned}
\mathrm{JHCDCov}^2(X_1,\ldots,X_q\mid Z)=C_2\sum_{i_1,i_2\in I_2^q}\mathrm{HCDCov}^2(X_{i_1},X_{i_2}\mid Z)\\
+C_3\sum_{i_1,i_2,i_3\in I_3^q}\mathrm{HCDCov}^2(X_{i_1},X_{i_2},X_{i_3}\mid Z)\\
+\cdots+C_q\,\mathrm{HCDCov}^2(X_1,\ldots,X_q\mid Z), \qquad \textbf{(2.4)}
\end{aligned}$$

for some specific constants $C_i>0$ with $2\le i\le q$.

JHCDCov is a linear combination of all HCDCov in $X_1,\ldots,X_q$ given Z with positive linear coefficients; hence, it can measure the conditional independence among $X_1,\ldots,X_q$ given Z.

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

PROPOSITION 2.2: $\mathrm{JHCDCov}^2(X_1,\ldots,X_q \mid Z)=0$ *if and only if* $\{X_1,\ldots,X_q\}$ *are conditionally independent given* Z.

Now, we discuss the selection of C_i. The rule is to make $\mathrm{JHCDCov}^2(X_1,\ldots,X_q \mid Z)$ as simple as possible. It turns out that $C_i=c^{q-i}$ in (2.4) with $c>0$ is an appropriate choice because we can get the following result.

PROPOSITION 2.3: *For any constant* $c>0$, *if* $C_i=c^{q-i}$, *we have*

$$\mathrm{JHCDCov}^2(X_1,\ldots,X_q \mid Z=z)=E\left[\prod_{i=1}^{q}(U_i(X_i,X_i',Z)+c) \mid Z=z,Z'=z\right]-c^q \quad \textbf{(2.5)}$$

Now, we consider the sample versions of HCDCov and JHCDCov. Suppose $W_n=\{X^k,Z^k\}_{k=1}^n$ are independent and identically distributed observations of $\{X,Z\}$, X_i^k is the kth observation of X_i, and Z^k is the kth observation of Z. Let $g_k(Z)=K_H(Z-Z^k)$ and $g(Z)=\sum_{k=1}^n g_k(Z)$, where $K_H(\cdot)=1/\det(H)K(H^{-1}\cdot)$ with the Gaussian kernel function $K(\cdot)$. Here, H is the bandwidth matrix, whose determinant and inverse are $\det(H)$ and H^{-1}, respectively. The kernel estimator of the density function of Z is $g(Z)/n$. The estimator of $\phi_{X_i|Z}(t_i)$ is defined by

$$\phi^n_{X_i|Z}(t_i)=\frac{1}{g(Z)}\sum_{k=1}^{n}\exp(\imath\langle t_i,X_i^k\rangle)g_k(Z)$$

For $1\le k,l\le n$, let

$$\begin{aligned}\widehat{U}_i(k,l,Z)=&\frac{1}{g(Z)}\sum_{v=1}^{n}|X_i^k-X_i^v|_{p_i}g_v(Z)+\frac{1}{g(Z)}\sum_{u=1}^{n}|X_i^u-X_i^l|_{p_i}g_u(Z)\\&-|X_i^k-X_i^l|_{p_i}-\frac{1}{g^2(Z)}\sum_{u,v=1}^{n}|X_i^u-X_i^v|_{p_i}g_v(Z)g_u(Z)\end{aligned} \quad \textbf{(2.6)}$$

DEFINITION 2.3: *An empirical estimator of* $\mathrm{HCDCov}^2(X_1,\ldots,X_q \mid Z)$ *is defined as*

$$\begin{aligned}\mathrm{HCDCov}_n^2(W_n|Z)&=\mathrm{HCDCov}_n^2(X_1,\ldots,X_q \mid Z)\\&=\int_{R^p}\left|\frac{1}{g(Z)}\sum_{k=1}^{n}\left\{\prod_{i=1}^{q}\left[\phi^n_{X_i^k|Z}(t_i)-\exp(\imath\langle t_i,X_i^k\rangle)\right]\right\}g_k(Z)\right|^2 d\omega\end{aligned} \quad \textbf{(2.7)}$$

Furthermore, we have

$$\begin{aligned}\mathrm{JHCDCov}_n^2(X_1,\ldots,X_q \mid Z)=&\,c^{q-2}\sum_{i_1,i_2\in I_2^q}\mathrm{HCDCov}_n^2(X_{i_1},X_{i_2} \mid Z)\\&+c^{q-3}\sum_{i_1,i_2,i_3\in I_2^q}\mathrm{HCDCov}_n^2(X_{i_1},X_{i_2},X_{i_3} \mid Z)\\&+\cdots+\mathrm{HCDCov}_n^2(X_1,\ldots,X_q \mid Z)\end{aligned} \quad \textbf{(2.8)}$$

When $q=2$, $\mathrm{HCDCov}_n^2(W_n|Z)$ is $\mathrm{CDCov}_n^2(W_n|Z)$ of [12].

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

According to (2.7) and (2.8), it is very difficult to compute empirical estimators because multiple integrals are involved. However, the following Proposition 2.4 provides us simple plug-in estimators.

PROPOSITION 2.4:

$$\mathrm{HCDCov}_n^2(W_n|Z) = \frac{1}{g^2(Z)} \sum_{k,l=1}^{n} \prod_{i=1}^{q} \widehat{U}_i(k,l,Z) g_k(Z) g_l(Z) \tag{2.9}$$

Furthermore, we have

$$\mathrm{JHCDCov}_n^2(W_n|Z) = \frac{1}{g^2(Z)} \sum_{k,l=1}^{n} \prod_{i=1}^{q} (\widehat{U}_i(k,l,Z) + c) g_k(Z) g_l(Z) - c^q \tag{2.10}$$

REMARK 2.1: *When* $q = 2$, $\mathrm{HCDCov}_n^2(W_n|Z) = D_1 + D_2 - 2D_3$, *where*

$$D_1 = \frac{1}{g^2(Z)} \sum_{k,l=1}^{n} \prod_{i=1}^{2} |X_i^k - X_i^l|_{p_i} g_k(Z) g_l(Z),$$

$$D_2 = \frac{1}{g^4(Z)} \prod_{i=1}^{2} \sum_{k,l=1}^{n} |X_i^k - X_i^l|_{p_i} g_k(Z) g_l(Z),$$

$$D_3 = \frac{1}{g^3(Z)} \sum_{k,l,m=1}^{n} |X_1^k - X_1^l|_{p_1} |X_2^k - X_2^m|_{p_2} g_k(Z) g_l(Z) g_m(Z)$$

REMARK 2.2: *Common methods to select bandwidth are the plug-in and cross-validation [1,2,4,11]. We use the plug-in method which is simple. On the other hand, we find that our estimators are insensitive to c from the simulations. In application, we suggest to use* $c = 0.5$.

Now, we establish the asymptotic properties of $\mathrm{HCDCov}_n^2(W_n|Z)$ and $\mathrm{JHCDCov}_n^2(W_n|Z)$.

THEOREM 2.1: *Suppose* $E(|X_i|_{p_i}|Z) > \infty$ *for any* $i \in \{1, 2, \ldots, q\}$, *and* Z *has a density function. Then,*

$$\mathrm{HCDCov}_n^2(W_n|Z) \xrightarrow{\text{a.s.}} \mathrm{HCDCov}^2(X_1, \ldots, X_q \mid Z),$$
$$\mathrm{JHCDCov}_n^2(W_n|Z) \xrightarrow{\text{a.s.}} \mathrm{JHCDCov}^2(X_1, \ldots, X_q \mid Z),$$

as $n \to \infty$, *where* $\xrightarrow{\text{a.s.}}$ *denotes the almost sure convergence.*

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

3. CONDITIONAL MUTUAL INDEPENDENT TESTING

In the section, we study conditional independence hypothesis testing. From the above, JHCDC can measure the conditional independence. Hence, we use the following parameter

$$T(X, Z) = E[\text{JHCDCov}^2(X_1, \ldots, X_q \mid Z)] \tag{3.1}$$

and its plug-in estimator

$$T_n(W_n) = \frac{1}{n}\sum_{m=1}^{n} \text{JHCDCov}_n^2(W_n | Z = Z^m) \tag{3.2}$$

to measure the conditional independence.

$T_n(W_n)$ is a symmetric statistic. It is possible to derive the asymptotic distribution. But in this paper, we use the local bootstrap method [12] to compute the empirical p-value. Here, it is given in detail. Let $W_n^* = \{(X_1^k, \ldots, X_{i-1}^k, X_i^{*k}, X_{i+1}^k, \ldots, X_q^k, Z^k)\}_{k=1}^n$ be a bootstrap sample. Here, X_i^{*k} is generated from the empirical conditional distribution:

$$\hat{F}_{X_i|Z=Z^k}(x) = \frac{\sum_{j=1}^n g_j(Z^k) I_{(-\infty, X_i^j]}(x)}{\sum_{j=1}^n g_j(Z^k)}, \tag{3.3}$$

where $x \in R^{p_i}$,

$$I_A(x) = \begin{cases} 1 & amp; \text{if } x \in A, \\ 0 & amp; \text{if } x \notin A, \end{cases}$$

and X_i is called a local bootstrap variable. In W_n^*, we only resample X_i^{*k}. All others are the same as the original observation. We will discuss how to choose X_i at the end this section. Based on W_n^*, we can obtain $T_n(W_n^*)$. In the same way, we can get $\{T_n^1(W_n^*) \ldots T_n^B(W_n^*)\}$, where B is the replication time. Therefore, the empirical p-value is

$$\hat{p} = B^{-1}\sum_{b=1}^{B} I_{[T_n(W_n), +\infty)}(T_n^b(W_n^*))$$

The selection of the local bootstrap variable is very important. One criterion is to minimize the empirical p-value. We can also use the following rules. (1) We select one dimensional X_i. (2) Among these one dimensional X_i's, we choose the one which has the lowest correlation with Z. (3) If the dimensions of all random vectors are bigger than one, we select the one among p components of X which has the lowest correlation with Z.

4. SIMULATION

In this section, we investigate the finite sample performance of the proposed method via the size and power (rejection probabilities). We choose small and moderate sample sizes $n = \{50, 100, 150, 200\}$ and $c = \{0.5, 2\}$. The number of samples is 1,000. The nominal level is 0.05. The number of bootstrap replications is set to be $B = 200$. Examples 4.1–4.3 are designed for empirical size, and others for empirical power. Examples 4.7 and 4.8 are used to evaluate the criterion of selecting the local bootstrap variable.

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

EXAMPLE 4.1: $(X_1, X_2, X_3, Z)^\top$ *comes from the multivariate normal distribution with zero mean vector and the covariance matrix*

$$\begin{bmatrix} 1 & 0.36 & 0.36 & 0.6 \\ 0.36 & 1 & 0.36 & 0.6 \\ 0.36 & 0.36 & 1 & 0.6 \\ 0.6 & 0.6 & 0.6 & 1 \end{bmatrix}$$

and $X = (X_1, X_2, X_3)^\top$. X_1 *is the local bootstrap variable.*

EXAMPLE 4.2: V_1, V_2, V_3 *and* Z *are independent and identically distributed random variables from the binomial distribution* $\mathrm{Bin}(10, 0.5)$. *Let* $X_i = V_i + Z,\ i = 1, 2, 3$ *and* $X = (X_1, X_2, X_3)^\top$. X_1 *is the local bootstrap variable.*

EXAMPLE 4.3: $(X_1, X_2, X_3, Z_1, Z_2)^\top$ *comes from the multivariate normal distribution with zero mean vector and the covariance matrix*

$$\begin{bmatrix} 1 & 0.32 & 0.32 & 0.4 & 0.4 \\ 0.32 & 1 & 0.32 & 0.4 & 0.4 \\ 0.32 & 0.32 & 1 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1 & 0 \\ 0.4 & 0.4 & 0.4 & 0 & 1 \end{bmatrix},$$

$X = (X_1, X_2, X_3)^\top$ *and* $Z = (Z_1, Z_2)^\top$. X_1 *is the local bootstrap variable.*

EXAMPLE 4.4: $(X_1, X_2, X_3, Z)^\top$ *comes from the multivariate normal distribution with zero mean vector and the covariance matrix*

$$\begin{bmatrix} 1 & 0.7 & 0.7 & 0.6 \\ 0.7 & 1 & 0.7 & 0.6 \\ 0.7 & 0.7 & 1 & 0.6 \\ 0.6 & 0.6 & 0.6 & 1 \end{bmatrix}$$

and $X = (X_1, X_2, X_3)^\top$. *Furthermore, we get the conditional covariance matrix of* X *given* Z *following*

$$\begin{bmatrix} 0.64 & 0.34 & 0.34 \\ 0.34 & 0.64 & 0.34 \\ 0.34 & 0.34 & 0.64 \end{bmatrix}.$$

X_1 *is the local bootstrap variable.*

EXAMPLE 4.5: V *and* Z *are independent and identically distributed random variables from the binomial distribution* $\mathrm{Bin}(10, 0.5)$. *Let* $X_1 = V + Z, X_2 = (V - 5)^4 + Z,\ X_3 = (V - 10)^4 + Z$ *and* $X = (X_1, X_2, X_3)^\top$. X_1 *is the local bootstrap variable.*

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

TABLE 1. Empirical size or power of simulation studies

n	$c = 0.5$	$c = 2$	$c = 0.5$	$c = 2$	$c = 0.5$	$c = 2$
	Example 4.1		Example 4.2		Example 4.3	
50	0.063	0.053	0.049	0.060	0.073	0.067
100	0.054	0.051	0.035	0.052	0.056	0.066
150	0.042	0.036	0.039	0.049	0.055	0.044
200	0.044	0.056	0.047	0.052	0.052	0.055
	Example 4.4		Example 4.5		Example 4.6	
50	0.845	0.841	0.818	0.852	0.643	0.648
100	0.959	0.962	0.975	0.963	0.805	0.795
150	0.974	0.988	0.993	0.995	0.839	0.845
200	0.993	0.993	0.998	0.996	0.875	0.881
	Example 4.7-1		Example 4.7-2		Example 4.7-3	
50	0.680	0.712	0.638	0.623	0.952	0.954
100	0.905	0.893	0.923	0.925	0.996	0.995
150	0.964	0.975	0.976	0.978	0.999	0.998
200	0.991	0.988	0.990	0.988	0.999	1.000
	Example 4.8-1		Example 4.8-2		Example 4.8-3	
50	0.686	0.686	0.743	0.750	0.994	0.995
100	0.892	0.894	0.916	0.902	0.998	1.000
150	0.953	0.958	0.962	0.967	1.000	0.999
200	0.978	0.978	0.982	0.979	1.000	1.000

EXAMPLE 4.6: *$(X_1, X_2, X_3, Z_1, Z_2)^\top$ comes from the multivariate normal distribution with zero mean vector and the covariance matrix*

$$\begin{bmatrix} 1 & 0.9 & 0.9 & 0.4 & 0.4 \\ 0.9 & 1 & 0.9 & 0.4 & 0.4 \\ 0.9 & 0.9 & 1 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1 & 0 \\ 0.4 & 0.4 & 0.4 & 0 & 1 \end{bmatrix},$$

$X = (X_1, X_2, X_3)^\top$ and $Z = (Z_1, Z_2)^\top$. X_1 is the local bootstrap variable. Furthermore, we get the conditional covariance matrix of X given Z following

$$\begin{bmatrix} 0.68 & 0.58 & 0.58 \\ 0.58 & 0.68 & 0.58 \\ 0.58 & 0.58 & 0.68 \end{bmatrix}$$

EXAMPLE 4.7: *$V_1, \ldots, V_7$ are independent and identically distributed random variables from the Student's t distribution with degree 15. We define $X_1 = (V_1, V_2, V_3, U)^\top$, $U = V_4 + V_5$, $X_2 = V_1 + \cdots + V_6$ and $X_3 = V_1 + \cdots + V_5 + V_7$. Further, $X = (X_1, X_2, X_3)^\top$ and $Z = V_5$. Let V_1, U and X_3 be the respective local bootstrap variables. The corresponding results are presented in Table 1 with respective titles "Examples 4.7-1–4.7-3".*

EXAMPLE 4.8: *$(X_1, X_2, X_3, Z)^\top$ comes from the multivariate normal distribution with zero mean vector and the covariance matrix*

$$\begin{bmatrix} 1 & 0.7 & 0.7 & 0.8 \\ 0.7 & 1 & 0.7 & 0.6 \\ 0.7 & 0.7 & 1 & 0.3 \\ 0.8 & 0.6 & 0.3 & 1 \end{bmatrix}$$

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

TABLE 2. Distances for five cities (km)

	BJ	TJ	JN	TS	QHD
BJ	0	113	365	156	273
TJ	113	0	273	156	224
JN	365	273	0	344	427
TS	156	156	344	0	125
QHD	273	224	427	125	0

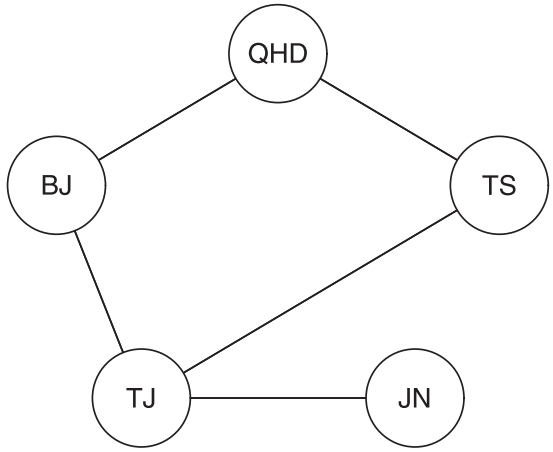


FIGURE 1. Gaussian graphical model for $PM_{2.5}$ data.

and $X = (X_1, X_2, X_3)^\top$. *Furthermore, we get the conditional covariance matrix of* X *given* Z *following*

$$\begin{bmatrix} 0.36 & 0.22 & 0.46 \\ 0.22 & 0.64 & 0.52 \\ 0.46 & 0.52 & 0.91 \end{bmatrix}$$

Let X_1, X_2 *and* X_3 *be the respective local bootstrap variables whose results correspond to Examples 4.8-1–4.8-3 in Table 1.*

Table 1 reports the results, which can be summarized as follows.

1. Empirical size of the proposed test is very close to 0.05, which supports the proposed method.
2. Empirical power of the proposed test increases as the sample size increases and tends to 1 quickly. The speed to 1 of power for one dimensional Z is faster than high dimensions because multivariate kernel density estimator needs more observations to achieve the same precision.
3. In Example 4.7, V_1 and U are components of X_1, X_3 is one dimensional. So it is expected that "Example 4.7-3" contains the best result. In Example 4.8, X_1, X_2, X_3 are one dimensional. But X_3 has the lowest correlation with Z. So "Examples 4.8-3" contains the best result. Actually the results of Examples 4.7 and 4.8 are consistent with the criteria for selecting the local bootstrap variable. Especially in small sample

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

size, Examples 4.7-3 and 4.8-3 are better than other cases, respectively. When the sample size is big, their powers are very high and are close to 1.

5. A REAL DATA

Now, we apply our proposed method to analyze the relationships of $PM_{2.5}$ in five Chinese cities by the Gaussian graphical model. The undirected graph $G = (V, E)$ where the nodes $V = \{V_k\}_{k=1}^5 = \{\text{BJ}, \text{TJ}, \text{JN}, \text{TS}, \text{QHD}\}$ represent the set of variables and E is a set of undirected edges. We use the notation X_N to denote the random variable corresponding node N. There is an edge between nodes V_i and V_j if and only if the random variables X_{V_i} and X_{V_j} are conditionally dependent, given all remaining variables for $1 \leq i \neq j \leq 5$. From Table 2 and Figure 1, distance is an important factor for $PM_{2.5}$ since closer cities are connected by edges. This is very reasonable because $PM_{2.5}$ has a strong regional performance. Since BJ, TS and JN are isolated after removing TJ and QHD in the figure, we test conditional independence of $X_{\text{BJ}}, X_{\text{TS}}$ and X_{JN} given X_{TJ} and X_{QHD} to justify the Gaussian graphical model. We get that the p-value of the proposed method is 0.255. If the nominal level is 0.05, $X_{\text{BJ}}, X_{\text{TS}}$ and X_{JN} are conditionally independent given X_{TJ} and X_{QHD}. This implies that TJ and QHD influence $PM_{2.5}$ of the other three cities to a large extent.

Acknowledgments

The authors thank the Editor, one Editorial member and one referee for their constructive suggestions that helped to improve the paper considerably. Pengfei Liu's research was supported by the National Natural Science Foundation of China (11501261 and 11671178), the State Scholarship funded by China Scholarship Council (CSC 201808320107), the Project of Natural Science Research in Jiangsu Province (15KJB110007), Qing Lan Project and the Project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions. Xuejun Ma's research was supported by the Natural Science Foundation of Jiangsu Province (Grant No. SBK2020040632). Wang Zhou's research was partially supported by a grant R-155-000-204-114 at the National University of Singapore.

References

1. Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* 71(2): 353–360.
2. Chacon, J.E. & Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test* 19(2): 375–398.
3. Chakraborty, S. & Zhang, X. (2018). Distance metrics for measuring joint dependence with application to causal inference. *Journal of the American Statistical Association* 114(528): 1638–1650.
4. Duong, T. & Hazelton, M.L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics* 32(3): 485–506.
5. Edwards, D (2012). *Introduction to graphical modelling*. New York: Springer.
6. Huang, T.M. (2010). Testing conditional independence using maximal nonlinear conditional correlation. *The Annals of Statistics* 38(4): 2047–2091.
7. Linton, O. & Gozalo, P. (2014). Testing conditional independence restrictions. *Econometric Reviews* 33(5–6): 523–552.
8. Scutari, M. (2009). Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software* 35(3): 1–22.
9. Su, L. & White, H. (2014). Testing conditional independence via empirical likelihood. *Journal of Econometrics* 182(1): 27–44.
10. Szekely, G.J., Rizzo, M.L., & Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* 35(6): 2769–2794.
11. Wand, M.P. & Jones, M.C. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics* 9(2): 97–116.
12. Wang, X., Pan, W., Hu, W., Tian, Y., & Zhang, H. (2015). Conditional distance correlation. *Journal of the American Statistical Association* 110(512): 1726–1734.

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

APPENDIX

LEMMA 1: *For complex number $y_1, \ldots, y_n$ with $n \geq 2$, then for any $r > 1$*

$$\left| \sum_{i=1}^{n} y_i \right|^r \leq n^{r-1} \sum_{i=1}^{n} |y_i|^r$$

The lemma can be found in [3].

PROOF OF PROPOSITION 2.1.:

$$\begin{aligned}
&\int_{R^{p_i}} \{\phi_{X_i|Z=z}(t)\overline{\phi_{X_i|Z=z}(t)} - \phi_{X_i|Z=z}(t)\exp(-\imath\langle t, x'\rangle) \\
&\quad - \exp(\imath\langle t, x\rangle)\overline{\phi_{X_i|Z=z}(t)} + \exp(\imath\langle t, x\rangle)\exp(-\imath\langle t, x'\rangle)\} d\omega_{p_i} \\
&\quad = \int_{R^{p_i}} \{E[\exp(\imath\langle t, X_i - X_i'\rangle) \mid Z = z, Z' = z] - E[\exp(\imath\langle t, X_i - x'\rangle) \mid Z = z] \\
&\qquad - E[\exp(\imath\langle t, x - X_i'\rangle) \mid Z' = z] + \exp(\imath\langle t, x - x'\rangle)\}\, d\omega_{p_i} \\
&\quad = \int_{R^{p_i}} E\{[\exp(\imath\langle t, X_i - X_i'\rangle) \mid Z = z, Z' = z] - 1 + 1 - [\exp(\imath\langle t, X_i - x'\rangle) \mid Z = z] \\
&\qquad + 1 - [\exp(\imath\langle t, x - X_i'\rangle) \mid Z' = z] - (1 - \exp(\imath\langle t, x - x'\rangle))\} d\omega_{p_i} \\
&\quad = -E(|X_i - X_i'|_{p_i} \mid Z = z, Z' = z) + E(|X_i' - x|_{p_i} \mid Z = z) \\
&\qquad + E(|x - X_i'|_{p_i} \mid Z' = z) - |x - x'|_{p_i}, \\
&\quad = U_i(x, x', Z)
\end{aligned}$$

The third identity is from the Lemma 1 of [10]. ■

PROOF OF PROPOSITION 2.2.: According to the definition of JHCDCov, if $\text{JHCDCov}^2(X_1, \ldots, X_q \mid Z) = 0$, $\text{HCDCov}^2(X_{i_1}, \ldots X_{i_k} \mid Z) = 0, k = 2, \ldots, q.$. Hence, we can get $X_1, \ldots X_q$ are independent. The other direction is trivial. ■

PROOF OF PROPOSITION 2.3.: Note that

$$\begin{aligned}
&\prod_{i=1}^{q} (U_i(X_i, X_i', Z) + c) \\
&\quad = c^q + c^{q-1} \sum_{i=1}^{q} U_i(X_i, X_i', Z) + c^{q-2} \sum_{i_1, i_2 \in I_2^q} U_{i_1}(X_{i_1}, X_{i_1}', Z) U_{i_2}(X_{i_2}, X_{i_2}', Z) \\
&\qquad + \cdots + \prod_{i=1}^{q} U_i(X_i, X_i', Z)
\end{aligned}$$

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

Hence,

$$
\begin{aligned}
&E\left[\prod_{i=1}^{q}(U_i(X_i,X_i',Z)+c)|Z=z,Z'=z\right]-c^q\\
&\quad=c^{q-1}\sum_{i=1}^{q}E\left[U_i(X_i,X_i',Z)|Z=z,Z'=z\right]\\
&\qquad+c^{q-2}\sum_{i_1,i_2\in I_2^q}E\left[U_{i_1}(X_{i_1},X_{i_1}',Z)U_{i_2}(X_{i_2},X_{i_2}',Z)|Z=z,Z'=z\right]\\
&\qquad+\cdots+E\left[\prod_{i=1}^{q}U_i(X_i,X_i',Z)|Z=z,Z'=z\right]\\
&\quad=c^{q-2}\sum_{i_1,i_2\in I_2^q}\mathrm{HCDCov}^2(X_{i_1},X_{i_2}|Z=z)+\cdots+\mathrm{HCDCov}^2(X_1,\ldots,X_q|Z=z)\\
&\quad=\mathrm{JHCDCov}^2(X_1,\ldots,X_q|Z=z)
\end{aligned}
$$

■

Proof of Proposition 2.4.: It suffices to prove (2.9) since (2.10) follows via the definition of JHCDCov. From the definition of HCDCov, we need to show that

$$
\int_{R^p}\left|\frac{1}{g(Z)}\sum_{k=1}^{n}\left\{\prod_{i=1}^{q}\left[\phi^n_{X_i^k|Z}(t_i)-\exp(\imath\langle t_i,X_i^k\rangle)\right]\right\}g_k(Z)\right|^2 d\omega=\frac{1}{g^2(Z)}\sum_{k,l=1}^{n}\prod_{i=1}^{q}\widehat{U}_i(k,l,Z)g_k(Z)g_l(Z) \tag{A.1}
$$

Note that

$$
\begin{aligned}
&\int_{R^p}\left|\frac{1}{g(Z)}\sum_{k=1}^{n}\left\{\prod_{i=1}^{q}\left[\phi^n_{X_i^k|Z}(t_i)-\exp(\imath\langle t_i,X_i^k\rangle)\right]\right\}g_k(Z)\right|^2 d\omega\\
&\quad=\int_{R^p}\frac{1}{g^2(Z)}\sum_{k,l=1}^{n}\left\{\prod_{i=1}^{q}\left[\phi^n_{X_i^k|Z}(t_i)-\exp(\imath\langle t_i,X_i^k\rangle)\right]\left[\overline{\phi^n_{X_i^k|Z}(t_i)}-\exp(-\imath\langle t_i,X_i^l\rangle)\right]\right\}g_k(Z)g_l(Z)\,d\omega\\
&\quad=\frac{1}{g^2(Z)}\sum_{k,l=1}^{n}\left\{\prod_{i=1}^{q}\int_{R^{p_i}}\left[\phi^n_{X_i^k|Z}(t_i)-\exp(\imath\langle t_i,X_i^k\rangle)\right]\left[\overline{\phi^n_{X_i^k|Z}(t_i)}-\exp(-\imath\langle t_i,X_i^l\rangle)\right]d\omega_i\right\}g_k(Z)g_l(Z)
\end{aligned}
$$

and

$$
\begin{aligned}
&\left[\phi^n_{X_i|Z}(t_i)-\exp(\imath\langle t_i,X_i^k\rangle)\right]\left[\overline{\phi^n_{X_i|Z}(t_i)}-\exp(-\imath\langle t_i,X_i^l\rangle)\right]\\
&\quad=\frac{1}{g(Z)^2}\sum_{u,v=1}^{n}\exp(\imath\langle t_i,X_i^u-X_i^v\rangle)g_u(Z)g_v(Z)-\frac{1}{g(Z)}\sum_{u=1}^{n}\exp(\imath\langle t_i,X_i^u-X_i^l\rangle)g_u(Z)\\
&\qquad-\frac{1}{g(Z)}\sum_{v=1}^{n}\exp(\imath\langle t_i,X_i^k-X_i^v\rangle)g_v(Z)+\exp(\imath\langle t_i,X_i^k-X_i^l\rangle)\\
&\quad=-\frac{1}{g(Z)^2}\sum_{u,v=1}^{n}[1-\exp(\imath\langle t_i,X_i^u-X_i^v\rangle)]g_u(Z)g_v(Z)+\frac{1}{g(Z)}\sum_{u=1}^{n}[1-\exp(\imath\langle t_i,X_i^u-X_i^l\rangle)]g_u(Z)\\
&\qquad+\frac{1}{g(Z)}\sum_{v=1}^{n}[1-\exp(\imath\langle t_i,X_i^k-X_i^v\rangle)]g_k(Z)-[1-\exp(\imath\langle t_i,X_i^k-X_i^l\rangle)]
\end{aligned}
$$

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

From Proposition 2.1, we can obtain

$$\begin{aligned}
&\int_{R_i} -\frac{1}{g(Z)^2}\sum_{u,v=1}^{n}[1-\exp(\imath\langle t_i, X_i^u - X_i^v\rangle)]g_u(Z)g_v(Z) + \frac{1}{g(Z)}\sum_{u=1}^{n}[1-\exp(\imath\langle t_i, X_i^u - X_i^l\rangle)]g_u(Z) \\
&\quad + \frac{1}{g(Z)}\sum_{v=1}^{n}[1-\exp(\imath\langle t_i, X_i^k - X_i^v\rangle)]g_v(Z) - [1-\exp(\imath\langle t_i, X_i^k - X_i^l\rangle)]\, d\omega_i \\
&\quad = -\frac{1}{g(Z)^2}\sum_{u,v=1}^{n}\left\{\int_{R_i}[1-\exp(\imath\langle t_i, X_i^u - X_i^v\rangle)]d\omega_i\right\} g_u(Z)g_v(Z) \\
&\qquad + \frac{1}{g(Z)}\sum_{u=1}^{n}\left\{\int_{R_i}[1-\exp(\imath\langle t_i, X_i^u - X_i^l\rangle)]\, d\omega_i\right\} g_u(Z) \\
&\qquad + \frac{1}{g(Z)}\sum_{v=1}^{n}\left\{\int_{R_i}[1-\exp(\imath\langle t_i, X_i^k - X_i^v\rangle)]\, d\omega_i\right\} g_v(Z) \\
&\qquad - \int_{R_i}[1-\exp(\imath\langle t_i, X_i^k - X_i^l\rangle)]\, d\omega_i \\
&\quad = -\frac{1}{g(Z)^2}\sum_{u,v=1}^{n}|X_i^u - X_i^v|_{p_i} g_u(Z)g_v(Z) + \frac{1}{g(Z)}\sum_{u=1}^{n}|X_i^u - X_i^l|_{p_i} g_u(Z) \\
&\qquad + \frac{1}{g(Z)}\sum_{v=1}^{n}|X_i^k - X_i^v|_{p_i} g_v(Z) - |X_i^k - X_i^l|_{p_i} \\
&\quad = \widehat{U}_i(k,l,Z)
\end{aligned}$$

Hence, (A.2) can be proved. ■

Proof of Theorem 2.1.: Define

$$\xi_n(t_1,\ldots,t_q) = \frac{1}{g(Z)}\sum_{k=1}^{n}\left\{\prod_{i=1}^{q}\left[\phi^n_{X_i|Z}(t_i) - \exp(\imath\langle t_i, X_i^k\rangle)\right]\right\} g_k(Z)$$

So

$$\mathrm{HCDCov}_n^2(W_n|Z) = \int_{R^p}|\xi_n(t_1,\ldots,t_q)|^2\, d\omega$$

Let $u_k^i = \exp(\imath\langle t_i, X_i^k\rangle) - \phi_{X_i|Z}(t_i)$. Then,

$$\xi_n(t_1,\ldots,t_q) = \frac{1}{g(Z)}\sum_{k=1}^{n}\prod_{i=1}^{q}\left(\frac{1}{g(Z)}\sum_{j=1}^{n}u_j^i g_j(Z) - u_k^i\right) g_k(Z)$$

Since $|u_k^i| \le 2$,

$$|\xi_n(t_1,\ldots,t_q)| \le 2^q$$

For given $\delta > 0$ define the region

$$D(\delta) = \{(t_1,\ldots,t_q) : \delta \le |t_i|_{p_i} \le 1/\delta,\ i = 1,\ldots,1\}$$

and random variable

$$\mathrm{HCDCov}_{n,\delta}^2 = \int_{D(\delta)}|\xi_n(t_1,\ldots,t_q)|^2\, d\omega \le \int_{D(\delta)} 4^q\, d\omega > \infty$$

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

According to the Lebesgue dominated convergence theorem, for each δ, it follows that almost surely

$$\lim_{n\to\infty} \mathrm{HCDCov}^2_{n,\delta} = \mathrm{HCDCov}^2_{\delta} = \int_{D(\delta)} \left| E\left[\prod_{i=1}^{q}\left(\phi_{X_i|Z}(t_i) - \exp(\imath\langle t_i, X_i\rangle)\right) Z\right]\right|^2 d\omega$$

Clearly, $\mathrm{HCDCov}^2_{\delta}$ converges to HCDCov^2 as δ tends to zero. Now, it remains to prove that almost surely

$$\limsup_{\delta\to 0}\limsup_{n\to\infty} |\mathrm{HCDCov}^2_{n,\delta} - \mathrm{HCDCov}^2_{n}| = 0 \tag{A.2}$$

Now, we write $D^c(\delta) = \bigcup_{i=1}^{q}(A_i^1 \cup A_i^2)$, where $A_i^1 = \{|t_i|_{p_i} < \delta\}$ and $A_i^2 = \{|t_i|_{p_i} > 1/\delta\}$ for $i = 1, \ldots, q$.

Hence, we have

$$|\mathrm{HCDCov}^2_{n,\delta} - \mathrm{HCDCov}^2_{n}| = \int_{D^c(\delta)} |\xi_n(t_1, \ldots, t_q)|^2 \, d\omega \le \sum_{\substack{i=1,\ldots,q \\ m=1,2}} \int_{A_i^m} |\xi_n(t_1, \ldots, t_q)|^2 \, d\omega$$

Note that

$$\begin{aligned}\xi_n(t) &= \sum_{B\subset\{1,2,\ldots,q\}} \frac{1}{g(z)} \sum_{k=1}^{n} \left[\prod_{i\in B^c} \frac{1}{g(z)} \sum_{j=1}^{n} u_j^i g_j(z) \cdot \prod_{i\in B}(-u_k^i)\right] g_k(z) \\ &= \sum_{B\subset\{1,2,\ldots,q\}} \prod_{i\in B^c} \frac{1}{g(z)} \sum_{j=1}^{n} u_j^i g_j(z) \cdot \left(\frac{1}{g(z)} \sum_{k=1}^{n} \prod_{i\in B}(-u_k^i) g_k(z)\right),\end{aligned}$$

where $B^c = \{1, 2, \ldots, q\}/B$. Hence,

$$\begin{aligned}|\xi_n(t)|^2 &\le 2^q \sum_{B\subset\{1,2,\ldots,q\}} \left(\prod_{i\in B^c} \frac{1}{g(z)} \sum_{j=1}^{n} u_j^i g_j(z)\right)^2 \cdot \left(\frac{1}{g(z)} \sum_{k=1}^{n} \prod_{i\in B}(-u_k^i) g_k(z)\right)^2 \\ &=: 2^q \sum_{B\subset\{1,2,\ldots,q\}} \Delta_B\end{aligned}$$

Now, suppose $B = \{\gamma_1, \ldots, \gamma_d\}$ and $B^c = \{\gamma_{d+1}, \ldots, \gamma_q\}$.

$$\begin{aligned}\Delta_B &= \left(\frac{1}{g(Z)} \sum_{k=1}^{n} u_k^{\gamma_1} \cdots u_k^{\gamma_d} g_k(Z)\right)^2 \prod_{i\in B^c} \left(\frac{1}{g(Z)} \sum_{k=1}^{n} u_k^i g_k(Z)\right)^2 \\ &\le \frac{\sum_{k=1}^{n} |u_k^{\gamma_1}|^2 \cdots |u_k^{\gamma_d}|^2 g_k(Z)}{g(Z)} \prod_{i\in B^c} \frac{\sum_{k=1}^{n} |u_k^i|^2 g_k(Z)}{g(Z)}\end{aligned} \tag{A.3}$$

u_k^i has the following properties which will be used.

$$\int_{R^{p_i}} \frac{|u_k^i|^2}{c_{p_i}|t_i|_{p_i}^{p_i+1}} \, dt_i \le 2(|X_i^k|_{p_i} + E(|X_i|_{p_i} \mid Z)), \tag{A.4}$$

$$\int_{|t_i|_{p_i}>\delta} \frac{|u_k^i|^2}{c_{p_i}|t_i|_{p_i}^{p_i+1}} \, dt_i \le 2E_{X_i}(|X_i^k - X_i|_{p_i} G(|X_i^k - X_i|_{p_i}\delta) \mid |Z), \tag{A.5}$$

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

$$\int_{|t_i|_{p_i}<1/\delta} \frac{|u_k^i|^2}{c_{p_i}|t_i|_{p_i}^{p_i+1}}\,dt_i \le 4\delta, \tag{A.6}$$

where E_{X_i} means the expectation with respect to X_i, and for $z=(z_1,\ldots,z_d)\in R^d$

$$G(y)=\int_{|z|_d>y}\frac{1-\cos z_1}{|z|_d^{1+d}}\,dz,$$

which is bounded by c_d and $\lim_{y\to 0}G(y)=0$. The derivations follows the proof of Theorem 1 of [12].

Now, we consider integration of Δ_B in $D^c(\delta)$. Given i, we discuss this integration in two cases.

Case 1: $j\in B$. Without loss of generality, we suppose $\gamma_1=j$. Firstly, now we consider A_j^1. With (A.4)–(A.6), we have

$$\begin{aligned}
&\int_{A_j^1}\Delta_B\,d\omega\\
&\le \int_{A_{\gamma_1}^1}\frac{\sum_{k=1}^n|u_k^{\gamma_1}|^2\ldots|u_k^{\gamma_d}|^2g_k(Z)}{g(Z)}\prod_{i\in B^c}\frac{\sum_{k=1}^n|u_k^i|^2g_k(Z)}{g(Z)}\,d\omega\\
&=\frac{1}{g(Z)}\sum_{k=1}^n\left[\int_{A_{\gamma_1}^1}|u_k^{\gamma_1}|^2\,d\omega_{\gamma_1}\int_{R^{p\gamma_2}}|u_k^{\gamma_2}|^2\,d\omega_{\gamma_2}\ldots\int_{R^{p\gamma_d}}|u_k^{\gamma_d}|^2\,d\omega_{\gamma_d}\right]g_k(Z)\\
&\quad\times\prod_{i\in B^c}\frac{\sum_{k=1}^n\int_{R^{p_i}}|u_k^i|^2d\omega_i g_k(Z)}{g(Z)}\\
&\le\frac{1}{g(Z)}\sum_{k=1}^n[2E_{X_{\gamma_1}}(|X_{\gamma_1}^k-X_{\gamma_1}|_{p_{\gamma_1}}G(|X_{\gamma_1}^k-X_{\gamma_1}|_{p_{\gamma_1}}\delta)\mid Z)\times 2(|X_{\gamma_2}^k|_{p_{\gamma_2}}+E(|X_{\gamma_2}|_{p_{\gamma_2}}\mid Z))\\
&\quad\times\cdots\times 2(|X_{\gamma_d}^k|_{p_{\gamma_d}}+E(|X_{\gamma_d}|_{p_{\gamma_d}}\mid Z))]g_k(Z)\times\prod_{i\in B^c}\frac{\sum_{k=1}^n 2(|X_{\gamma_i}^k|_{p_{\gamma_i}}+E(|X_{\gamma_i}|_{p_{\gamma_i}}\mid Z))g_k(Z)}{g(Z)}
\end{aligned}\tag{A.7}$$

Since $g(Z)/n$ is a consistent density function estimator of Z,

$$\limsup_{n\to\infty}\int_{A_j^1}\Delta_B\,d\omega\le 2^{2q-1}E_{X_{\gamma_1}}(|X_{\gamma_1}-X'_{\gamma_1}|_{p_{\gamma_1}}G(|X_{\gamma_1}-X'_{\gamma_1}|_{p_{\gamma_1}}\delta)\mid Z)\prod_{i\ne 1}E(|X_{\gamma_i}|_{p_{\gamma_i}}\mid Z))$$

Hence, it most surely,

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\int_{A_j^1}\Delta_B\,d\omega=0 \tag{A.8}$$

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

Now, we consider A_j^2

$$\begin{aligned}
&\int_{A_j^1} \Delta_B \, d\omega \\
&\leq \int_{A_{\gamma_1}^2} \frac{\sum_{k=1}^n |u_k^{\gamma_1}|^2 \ldots |u_k^{\gamma_d}|^2 g_k(Z)}{g(Z)} \prod_{i \in B^c} \frac{\sum_{k=1}^n |u_k^i|^2 g_k(Z)}{g(Z)} \, d\omega \\
&= \frac{1}{g(Z)} \sum_{k=1}^n \left[\int_{A_{\gamma_1}^2} |u_k^{\gamma_1}|^2 \, d\omega_{\gamma_1} \int_{R^{p\gamma_2}} |u_k^{\gamma_2}|^2 \, d\omega_{\gamma_2} \ldots \int_{R^{p\gamma_d}} |u_k^{\gamma_d}|^2 \, d\omega_{\gamma_d} \right] g_k(Z) \\
&\quad \times \prod_{i \in B^c} \frac{\sum_{k=1}^n \int_{R^{p_i}} |u_k^i|^2 \, d\omega_i g_k(Z)}{g(Z)} \\
&= \frac{1}{g(Z)} \left[4\delta \times 2(|X_{\gamma_2}^k|_{p_{\gamma_2}} + E(|X_{\gamma_2}|_{p_{\gamma_2}} \mid Z)) \times \cdots \times 2(|X_{\gamma_d}^k|_{p_{\gamma_d}} + E(|X_{\gamma_d}|_{p_{\gamma_d}} \mid Z)) \right] g_k(Z) \\
&\quad \times \prod_{i \in B^c} \frac{\sum_{k=1}^n 2(|X_{\gamma_i}^k|_{p_{\gamma_i}} + E(|X_{\gamma_i}|_{p_{\gamma_i}} \mid Z)) g_k(Z)}{g(Z)}
\end{aligned}$$

Similarly,

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \int_{A_j^2} \Delta_B \, d\omega = 0 \tag{A.9}$$

Case 2: $j \notin B$. Following the similar steps as case 1, we can prove

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \int_{A_j^2} \Delta_B \, d\omega = 0, \tag{A.10}$$

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \int_{A_j^1} \Delta_B \, d\omega = 0. \tag{A.11}$$

Note that for $B = \emptyset$ or $B = \{1, 2, \ldots, q\}$,

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \int_{A_j^2} \prod_{i=1}^q \left(\frac{1}{g(Z)} \sum_{j=1}^n u_j^i g_j(Z) \right)^2 d\omega = 0, \tag{A.12}$$

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \int_{A_j^1} \prod_{i=1}^q \left(\frac{1}{g(Z)} \sum_{j=1}^n u_j^i g_j(Z) \right)^2 d\omega = 0, \tag{A.13}$$

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \int_{A_j^2} \left(\frac{1}{g(Z)} \sum_{k=1}^n u_k^1 u_k^2 \ldots u_k^q g_k(Z) \right)^2 d\omega = 0, \tag{A.14}$$

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \int_{A_j^1} \left(\frac{1}{g(Z)} \sum_{k=1}^n u_k^1 u_k^2 \ldots u_k^q g_k(Z) \right)^2 d\omega = 0. \tag{A.15}$$

From (A.9)–(A.15), we can prove (A.2). So,

$$\mathrm{HCDCov}_n^2(W_n|Z) \xrightarrow{\text{a.s.}} \mathrm{HCDCov}^2(X_1, \ldots, X_q \mid Z)$$

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press

With the definition of $\mathrm{JHCDCov}_n^2(W_n|Z)$, we can get

$$\mathrm{JHCDCov}_n^2(W_n|Z) \xrightarrow{\text{a.s.}} \mathrm{JHCDCov}^2(X_1, \ldots, X_q \mid Z).$$

■

https://doi.org/10.1017/S026996482000039X Published online by Cambridge University Press